RNA-Protein Crosslinking Identifies Novel Targets For The Nuclear RNA Surveillance Machinery

Wiebke Wlotzka

A thesis presented for the degree of

Doctor of Philosophy

The University of Edinburgh

2011

Declaration

I thereby declare that I alone have composed this thesis and that, except where stated otherwise, the work presented herein is my own.

Wiebke Wlotzka,

March 2011

Acknowledgements

Working in David's lab was a highly beneficial experience, being surrounded by many knowledgeable and skilled scientists. I was able to learn so many different views and approaches to address problems from everybody in the lab. Many of the past and present members, too many to name them all, contributed to my success with advice and helpful discussions or providing reagents. Grzegorz Kudla supported this project by designing tools, performing the complete bioinformatics analyses and trying to teach me a little about it. A particular big thanks goes to Claudia Schneider for her constant interest in the project, endless discussions about science and career planning and how to carefully design and execute experiments; I have learnt so much.

I am most grateful to David for four years of guidance and support. He gave me the opportunity and the space to develop my own interests without ever loosing track. His enthusiasm about science seems endless and it was a pleasure to work with and to learn from him.

Finally I want thank my parents for their unconditional love and support. They never questioned my aims and ambitions no matter how far away from home I had to go. I think they will be very happy to hold this thesis in their hands. Danke.

Abbreviations

asRNA	antisense RNA
ChIP	Chromatin immuno precipitation
CID	CTD interaction domain
CRAC	Crosslinking and analyses of cDNAs
CTD	C-terminal domain (of RNA Pol II)
CUT	cryptic unstable transcript
Da	Dalton
DNA	Desoxyribonucleic acid
EDTA	Ethylene-diamine-tetraacetate
HTP	His-TEV-ProteinA
IgG	Immunoglobulin G
IP	Immuno precipitation
mRNA	messenger RNA
ncRNA	non-coding RNA
nt	Nucleotide
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
RNA	Ribonucleic acid
RNase	Ribonuclease
RNP	Ribonucleoprotein particle
rRNA	ribosomal RNA
RRM	RNA recognition motif
RT	Room temperature
RT-PCR	Reverse transcription PCR
SD	Synthetic dropout
SDS	Sodium dodexyl sulphate
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
SUT	Stable unannotated transcript

TAP	Tandem affinity purification
TBE	Tris-borate-EDTA
TBS	Tris buffered saline
TEV	Tobacco etch virus
tRNA	transfer RNA
TRAMP	Trf-Air-Mtr4 polyadenylation complex
UV	Ultra violet
WT	Wild type

Abstract

The RNA binding proteins Nrd1 and Nab3 function in transcription termination by RNA Pol II, acting via interactions with the CTD of the largest polymerase subunit, particularly on snRNA and snoRNA genes. They also participate in nuclear RNA surveillance and ncRNA degradation, functioning together with the exosome and the Trf-Air-Mtr4 polyadenylation (TRAMP) complexes. To better understand the signals for surveillance and ncRNA degradation, I applied an RNA-protein crosslinking approach in combination with Solexa sequencing. This approach identified *in vivo* binding sites for Nrd1, Nab3 and Trf4.

Several million sequences were recovered and mapped to the yeast genome. This identified three classes of substrates: 1) Expected targets, including snRNAs, snoRNAs and characterized short ncRNAs. 2) Unknown but anticipated substrates, including several hundred previously uncharacterized ncRNAs that lie antisense to protein coding genes (asRNAs). 3) Unexpected targets, including many Pol III transcribed precursor RNAs.

Bioinformatics analyses of the high-throughput sequencing data revealed that known binding motifs for Nrd1 and Nab3 were frequently recovered. Many recovered RNAs contained non-templated oligo(A) tails with an average of 2-5 nt length. This clearly distinguishes targets for surveillance machinery from polyadenylated mRNAs that get stabilized by polyadenylation ($A_{70\text{-}90}$ in yeast).

For a few selected, predicted asRNAs I was able to validate the crosslinking data by demonstrating that corresponding long RNAs are both detectable and increased by loss of Nrd1, Nab3, Trf4 or the exosome component Rrp6. Interestingly, loss of Nrd1 or Nab3 led to transcriptional read through on long asRNA transcripts. In addition, I have identified pre-*TLC1* (telomerase RNA) as a target for the surveillance machinery. Processing of this long ncRNA was only poorly characterized in yeast but I could demonstrate that its transcription termination and maturation is mainly dependent on actions of the Nrd1-Nab3-Sen1, TRAMP4 and exosome complexes. It

was previously reported that Nrd1-Nab3 acts only on short RNAs, due to the association with Ser5 phosphorylated CTD. My findings suggest that action of Nrd1-Nab3 is not exclusively on Ser5 phosphorylated form of the CTD.

Unexpectedly the Pol II associated factors Nrd1 and Nab3 bound Pol III precursor transcripts. Surveillance of Pol III transcripts was dependent on Nrd1 and Nab3 since depletion of Nrd1 or Nab3 led to accumulation of pre-tRNAs. In addition, I could demonstrate that pre-RNase P RNA is oligoadenylated *in vivo*, which was dependent on Nrd1, Nab3 and Trf4. Together, my findings suggest a revised model of nuclear RNA surveillance in which Nrd1-Nab3 not only acts in co-transcriptional RNA recognition on Pol II transcripts but also post-transcriptionally on Pol III RNAs. The TRAMP complex is recruited to the defective RNA by the Nrd1-Nab3 complex, which remains associated with the RNA through the process of polyadenylation, until the exosome degrades the aberrant transcript.

Aim of the project

All newly synthesized RNAs undergo maturation to form functional RNAs and quality control pathways apparently monitor many or all steps of the processing pathways. The exosome is a major player in RNA surveillance and is assisted by various cofactors that identify or mark defective RNAs and target them for degradation by the surveillance machinery. In addition to its degradative role in RNA surveillance, the exosome participates in the precise trimming and 3' maturation of several stable ncRNAs. Regulation of exosome activity is believed to be mediated by its cofactors, which may also help to distinguish between targets for the processing and degradation modes of exosome activity. Known cofactors include the TRAMP polyadenylation complex, RNA binding proteins Rrp47 and Mpp6 as well as the Pol II associated hnRNP proteins Nrd1 and Nab3. Polyadenylation of aberrant RNAs has been established as a signal for degradation in eukaryotes, as previously observed in bacteria. However, which properties make an RNA a target for polyadenylation and surveillance, and how these are recognized by the various factors, is not known. Few endogenous targets for the nuclear surveillance machinery have been identified, and even fewer have been characterized in any detail.

One aim of my PhD project was, therefore, to identify RNAs that are bound by factors of the surveillance machinery genome-wide in a wild type, unperturbed cell. This was particular exciting because with an unbiased approach, targets that had not been anticipated could potentially be found. Following the initial identification of targets, several examples would then be further characterized.

Furthermore I hoped to gain insights into which features the targeted RNAs had, that identified them as *bona fide* targets. Such target sites could reside within the RNA sequence, but not normally be detected; e.g. in intervening or flanking sequences that should be rapidly removed by processing, or masked by correct RNA folding or protein binding. Some factors are known to bind specific RNA motifs, this could be a trait that is common among surveillance factors. Another possibility is that structured RNAs that have not been able to adopt the correct conformation could be specifically recognized by surveillance factors.

Chapter 1

Introduction- almost everything RNA

1.1 The RNA surveillance machinery- Why is RNA quality control necessary?

Cells in exponential growth produce a large number of different RNAs. The most abundant of these are stable non-coding RNAs (ncRNA), including ribosomal RNAs (rRNA; >70% of all RNAs in the cell), transfer RNAs (tRNA) and small nuclear/nucleolar RNAs (sn(o)RNA). In addition, thousands of protein-coding messenger RNAs (mRNA) are transcribed, along with many unstable ncRNAs. Almost all RNAs undergo co- or post-transcriptional processing steps in order to form a functional molecule. RNA polymerase II (RNA Pol II) transcripts receive a 7-methyl guanosine cap at the 5' end while the RNA is emerging from the polymerase, which is further processed to a 2,2,7-trimethyl guanosine cap in case of snRNAs. Furthermore, mRNAs containing non-coding intronic sequences are spliced and all mRNAs get cleaved and polyadenylated at their 3'end. Stable ncRNAs generally obtain base modifications and in the case of rRNAs and human snRNAs, these are even guided by ribonucleoprotein particles (RNPs), which contain snoRNAs or small Cajal RNAs (scaRNAs), respectively. Most RNAs function as RNA-protein (RNP) complexes, which are assembled in complex, stepwise maturation pathways. Ribosomal RNA associates with many different proteins for processing and assembly in order to form a ribosome. The snRNAs associate with proteins to form spliceosomal snRNPs, which assemble on mRNAs to form catalytically active spliceosomes. Even mRNAs that are exported to the cytoplasm for translation associate with sets of proteins (heterogeneous nuclear proteins; forming an hnRNP), which promote packaging and nuclear export, and regulate translation and turnover in the cytoplasm. All of these processing steps that convert the primary transcript into a mature RNA/RNP have the potential to make mistakes. In order to guarantee high fidelity of gene expression, eukaryotic cells have developed active RNA quality control systems. This surveillance machinery monitors every step in the processing pathways that convert precursor RNAs into mature RNA-protein complexes. Several protein complexes are involved in monitoring RNA processing. These can recognise and mark aberrant RNAs (e.g. Nrd1-Nab3 and the TRAMP complex) and several exonucleases (including the exosome complex) are able to degrade defective RNAs.

Frequently, multiple nucleases are able to degrade the same target RNAs, making redundancy a general feature of RNA degradation systems. Such redundancy is predicted to ensure robust, high fidelity surveillance of gene expression.

1.1.1 The exosome (3' to 5' degradation)

The exosome is the major instrument of 3' to 5' RNA degradation in yeast, and probably in many other eukaryotes. It is a multisubunit complex with 3' exonuclease and endonuclease activity that degrades RNAs that have been targeted by surveillance systems in the nucleus or the cytoplasm (reviewed by Houseley et al., 2006; Houseley and Tollervey, 2009). In addition to its role in complete RNA degradation, the exosome complex is also engaged in the precise trimming of the 3' ends of several precursor RNA species (Houseley et al., 2006).

1.1.1.1 Subunits of the exosome complex

The eukaryotic core exosome consists of 10 distinct polypeptides that form a barrel like structure (Allmang et al., 1999b; Bonneau et al., 2009; Liu et al., 2006; Mitchell et al., 1997). Six subunits, Rrp41, Rrp42, Rrp43, Rrp45, Rrp46 and Mtr3 form a ring structure. The ring forming subunits are related to *E. coli* RNase PH, which is a phosphorolytic 3' to 5' exonuclease that utilizes inorganic phosphate as a nucleophile to cleave the RNA backbone and release ribonucleoside 5' diphosphate products (reviewed by Symmons et al., 2002). However, all six PH-related components of the yeast and human exosomes have apparently been catalytically inactivated by point mutations. Three additional subunits, Csl4, Rrp4 and Rrp40 form the 'lid' of the barrel. They contain S1 or KH domains and are therefore thought to bind RNA. Rrp44, the tenth subunit of the yeast exosome, is a processive, hydrolytic exonuclease, which is related to bacterial RNase II (Mitchell et al., 1997). *In vitro* assays with purified exosomes showed that in the yeast core exosome, Rrp44 is the only active exonuclease (Dziembowski et al., 2007; Liu et al., 2006). Recently genetic approaches and *in vivo* RNA analyses showed that some phenotypes of Rrp44 depletions could not be attributed to the lack of the exonuclease activity

(Schaeffer et al., 2008; Schneider et al., 2009). Mutations in the PIN domain of Rrp44 were proven to be responsible for these phenotypes. In *vitro* assays with purified Rrp44 demonstrated that the PIN domain harbours endoribonuclease activity. Furthermore it was shown that the PIN domain facilitates protein-protein interaction and mediates association of Rrp44 with the exosome ring (Schneider et al., 2009).

The structure of the eukaryotic exosome closely resembles the archaeal exosome and bacterial PNPase, and all 10 core exosome subunits are essential for viability (Houseley et al., 2006 and references therein). This raises the question why the structure of the exosome was conserved through evolution but not its activity? Recent structural investigations indicate that RNA substrates are 'fed' through the exosome core structure into the exonuclease domain of Rrp44 (Figure 1.1 adapted from Bonneau, 2009). Conti and colleagues demonstrated that the channel formed by the nine-subunit core exosome binds RNA substrates. The identified binding sites used for RNA recognition are similar to the substrate binding sites in the prokaryotic counterparts underlining the evolutionary conservation of this RNA degrading machine (marked nucleotides in the model in Figure 1.1). Furthermore they showed that Rrp44 associates with the exosome core by interaction with Rrp41-Rrp45 on the 'bottom' of the barrel through its PIN domain. From their data they infer that the 3' end of an RNA substrate is led by the S1/KH proteins through the central channel of the RNase PH-like ring to the exoribonucleolytic site of Rrp44 for degradation. In this scenario the catalytically inactive subunits could be responsible for regulating the activity of Rrp44 through protein-RNA and protein-protein interactions.

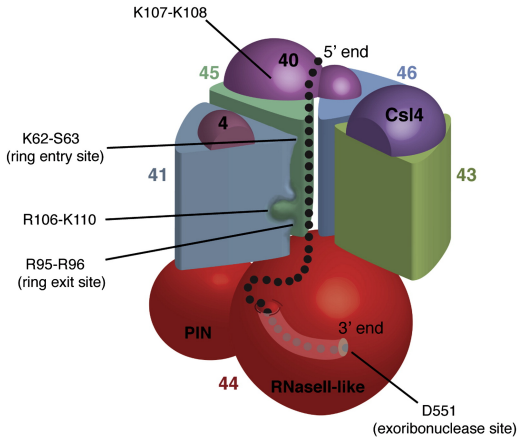

Figure 1.1 Schematic representation of the eukaryotic exosome and its RNA substrate (Bonneau et al., 2009)

The exosome complex is present in the nucleus and cytoplasm but the core components associate with distinct nuclear and cytoplasmic factors. In the cytoplasm the exosome is accompanied by the GTPase Ski7, which interacts also with the Ski2/3/8 complex composed of the putative RNA helicase Ski2, a tetratricopeptide repeat protein Ski3 and a WD repeat protein Ski8 (Brown et al., 2000). Together they participate in general mRNA turnover (Anderson and Parker, 1998; van Hoof et al., 2002) and in the surveillance of mRNAs with a premature stop codon (NMD) Mitchell and Tollervey, 2003; Takahashi et al., 2003) or mRNAs lacking a translation-termination codon (No-go decay; Frischmeyer et al., 2002; van Hoof et al., 2002).

In the nucleus, the exosome associates with Rrp6, a distributive, hydrolytic exonuclease that shares sequence similarity with E. coli RNase D (Burkard and Butler, 2000) and the non-essential, putative nucleic-acid-binding protein Rrp47 (Mitchell et al., 2003). Rrp6 is not essential for viability but cells lacking Rrp6 exhibit a temperature sensitive (ts) growth phenotype. Moreover the RNA processing/degradation phenotypes of *rrp6*Δ strains or point mutants are remarkably different from the phenotypes observed in core exosome mutants (Allmang et al., 1999a; Allmang et al., 1999b; Bousquet-Antonelli et al., 2000; Briggs et al., 1998; Mitchell et al., 2003). *In vitro* assays also showed clear differences in degradation

activities of Rrp6 containing exosomes (distributive) and Rrp44 exosomes (processive) (Liu et al., 2006), implying that RNA substrates of Rrp6 do not normally transit through the central channel of the exosome core.

1.1.1.2 Functions of the nuclear exosome

In the nucleus, the exosome complex is responsible for the surveillance and degradation of aberrant RNAs. These surveillance activities of the nuclear exosome encompass degradation of rRNA precursors (Allmang et al., 2000), or aberrant 5S rRNA (Kadaba et al., 2006), tRNA precursors (Copela et al., 2008; Kadaba et al., 2004; Kadaba et al., 2006; Vanacova et al., 2005), cryptic unstable transcripts (CUTs) (Wyers et al., 2005) as well as pre-mRNAs with defects in intron removal or 3' end formation (Bousquet-Antonelli et al., 2000; Hilleren et al., 2001; Milligan et al., 2005). A detailed description of the activities of the surveillance machinery on relevant substrates is presented in section 1.2.

On the other hand the exosome also functions in the precise processing of RNA precursors in the nucleus and was identified through its activity in 3' end formation of 5.8S rRNA (Mitchell et al., 1997). Furthermore, it is responsible for the generation of mature 3' ends of snoRNAs, snRNAs and a few mRNA species (Allmang et al., 1999a; Ciais et al., 2008; Roth et al., 2009). This role of the exosome complex in the precise trimming of certain RNA substrates stands in contrast to the complete degradation of surveillance targets and implies the need for tight regulation of its activity. Specific signals mark transcripts, to allow the exosome to distinguish which RNAs have to undergo processing by nucleotide removal and which are non-functional and need to be degraded. These signals presumably reside within the sequence and/or structure of the RNA and associated proteins. Processing factors that are still associated with an RNA that has not successfully undergone processing could serve as a signal, potentially similar to the function of the exon junction complex in human nonsense mediated RNA decay (NMD). Cofactors that regulate exosome activity and mark RNAs for degradation will be discussed in section 1.1.3.

1.1.2 5’ to 3’ degradation by the exonucleases Rat1 and Xrn1

Eukaryotic RNAs are degraded from the 5’end, if this is not protected - by a cap, 5’ triphosphate, secondary structure or proteins. The factors Rat1 and Xrn1 function as general 5’ to 3’ exoribonucleases in the nucleus and the cytoplasm, respectively, and are 39% similar/identical (Amberg et al., 1992; Johnson, 1997; Kenna et al., 1993; Larimer et al., 1992).

Rat1 is an essential protein and is associated with a cofactor Rai1, which supports its structure and therefore promotes its exoribonuclease activity (Xiang et al., 2009). In the nucleus the Rat1-Rai1 complex functions in maturation of the 5’ end of 5.8S rRNA, rRNA spacer fragments and polycystronic and intronic snoRNAs (Amberg et al., 1992; Petfalski et al., 1998). Rai1 is a pyrophophatase that removes 5’-triphosphates and defective cap structures allowing degradation by Rat1, which requires substrates with a 5’monophosphate group (Xiang et al., 2009).

In addition, Rat1-Rai1 functions in transcription termination of RNA Pol I and Pol II. In the torpedo model, Rat1 chases the elongating polymerase after the primary transcript has been cleaved, behind a stem-loop structure by Rnt1 in case of 35S pre-rRNA or after passing through the cleavage and polyadenylation signal in case of Pol II. These cleavage events generate an exposed and unprotected 5’-monophosphate end; degradation of the downstream cleavage product forces the polymerase to stop (El Hage et al., 2008; Kawauchi et al., 2008; Kim et al., 2004).

In the cytoplasm, the non-essential factor Xrn1 is involved in mRNA degradation and quality control following removal of the protective cap structure by the decapping complex Dcp1-Dcp2 (reviewed by Doma and Parker, 2007). Loss of the catalytic activities of both Rrp44 and Xrn1 generates a synthetic lethal interaction, indicating that these proteins carry the major cytoplasmic RNA degradation activities (Schneider et al., 2009).

1.1.3 Cofactors for the exosome

Both RNA degradation and RNA processing by the exosome require cofactors, which interact with specific exosome components (Milligan et al., 2008; Mitchell et al., 2003; Vasiljeva and Buratowski, 2006) and/or recognise RNA substrates and mark them for degradation (Kadaba et al., 2004; LaCava et al., 2005a; Vanacova et al., 2005; Wyers et al., 2005). These help the exosome to distinguish processing from surveillance targets and carry out the appropriate function as well as regulating its activity.

1.1.3.1 The TRAMP complex

Polyadenylation of RNA in bacteria was known to destabilize the target RNAs, whereas polyadenylation in eukaryotes was widely believed to only protect the RNA from degradation and promote export and translation (Houseley and Tollervey, 2009). This was challenged by discovery of the TRAMP (Trf-Air-Mtr4 polyadenylation) complex that has been characterised as an important cofactor for the nuclear exosome, as it marks aberrant RNAs with an oligo(A) tail and targets them for degradation (LaCava et al., 2005a; Vanacova et al., 2005).

In yeast, this oligoadenylation activity resides in Trf4 and Trf5 (57% identity), which are homologues of the nuclear poly(A) polymerase Pap1. Trf4 and Trf5 are dispensable for cell viability, however, *trf4*Δ strains are cold sensitive (cs) lethal with slow growth at all temperatures and double mutants are synthetically lethal, indicating redundant functions. Members of this family of non-canonical polymerases were initially identified in *S.pombe* (Cid1, Cid13) and *C.elegans* (Gld2) (Rouhana et al., 2005; Saitoh et al., 2002; Wang et al., 2000a). Their domain organisation is very similar to the canonical Pap1 (Figure 1.2) with an N-terminal catalytic domain including three conserved aspartate residues and a conserved strand loop motif followed by a central domain. They however lack the RNA binding domain that is present in Pap1 (reviewed by Martin et al., 2008).

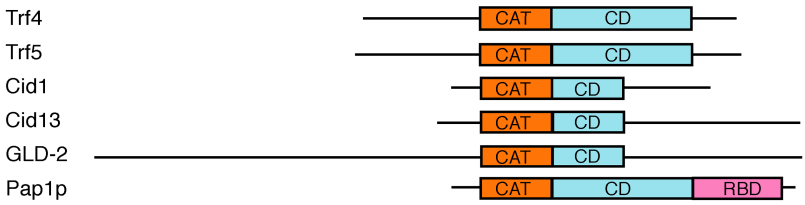

Figure 1.2 Poly(A) polymerase domain organization (Vanacova et al., 2005)

Trf4 and Trf5 proteins show overlapping but not fully redundant functions (Houseley and Tollervey, 2006). Trf4 was first discovered in a synthetic genetic interaction screen with topoisomerase I (Top1; Sadoff et al., 1995) and was ascribed a function in sister chromatid cohesion and as a DNA polymerase (Haracska et al., 2005; Wang et al., 2002; Wang et al., 2000b). However, further studies failed to confirm the DNA polymerase activity. Trf4 and Rrp44 were subsequently found to be suppressors of a *trm6* mutation that otherwise leads to degradation of the initiator tRNA ($tRNA^{iMet}$), due to the absence of a single nucleotide modification (Kadaba et al., 2004). This study showed that the defective tRNA was polyadenylated and degraded, and that Trf4 functions in this pathway together with Rrp44 and Rrp6. In a different study, Trf4 and Trf5 were found to interact with Mtr4 (as a bait) in a two-hybrid assay (LaCava et al., 2005a). Mtr4 is a DExH box helicase, which is required as a cofactor for all tested nuclear functions of the exosome (Bernstein et al., 2008; de la Cruz et al., 1998; Wang et al., 2008). Trf4 and Trf5 form separate complexes with the zinc-knuckle, RNA binding proteins Air1 or Air2, together with and Mtr4 (LaCava et al., 2005a; Vanacova et al., 2005). Target recognition by the TRAMP complex is presumably facilitated by Air 1 and Air2 *in vivo* and they stimulate polyadenylation activity of Trf4 *in vitro* (Vanacova et al., 2005). The helicase activity of Mtr4 was shown to be necessary for Rrp6 to degrade structured RNAs *in vitro* (Vanacova et al., 2005) and it probably also serves as a scaffold since it functionally interacts with both the TRAMP complex and the exosome. Genetic interactions between TRAMP components and chromatin remodelling factors link TRAMP activity to chromatin remodelling, most likely via ncRNA stability (reviewed by Houseley and Tollervey, 2008).

1.1.3.2 The hnRNP proteins Nrd1 and Nab3

Nrd1 and Nab3 are essential, nuclear RNA binding proteins, which form a stable complex and also interact with the putative ATP-dependent RNA helicase Sen1 (Conrad et al., 2000; Steinmetz and Brow, 1996; Wilson et al., 1994). Nrd1 and Nab3 contain several hnRNP-like features, including a single consensus RNA recognition motif (RRM) and a proline- and glutamine-rich C-terminal domain (P/Q domain). Nrd1 also contains a short segment rich in arginine, serine, and glutamate, which is similar to RE/RS domains found in many metazoan splicing factors that are rich in RE and RS dipeptides. Other Nrd1 domains interact with the C-terminal domain (CTD) of Pol II and with Nab3 (Conrad et al., 2000; Steinmetz and Brow, 1996; the domain organizations of Nrd1 and Nab3 are displayed in Figure 1.3). Some domains in Nrd1 are dispensable for viability but loss of the ability to interact with either the RNA or Nab3 is lethal (Conrad et al., 2000; Steinmetz and Brow, 1998; Vasiljeva et al., 2008a). Nrd1 was first identified as a factor that down-regulated the expression of a pre-mRNA reporter construct in a sequence specific manner and it was speculated that this reflected enhanced degradation (Steinmetz and Brow, 1996). Nab3 is a very acidic protein as it includes an N-terminal stretch of aspartic and glutamic residues (Wilson et al., 1994). It was first characterised as an hnRNP-like protein associated with poly(A)$^+$ mRNAs. Initially characterised *nab3* mutants exhibited defects in pre-mRNA splicing and mRNA stability but not on 3' end formation or mRNA export (Wilson et al., 1994).

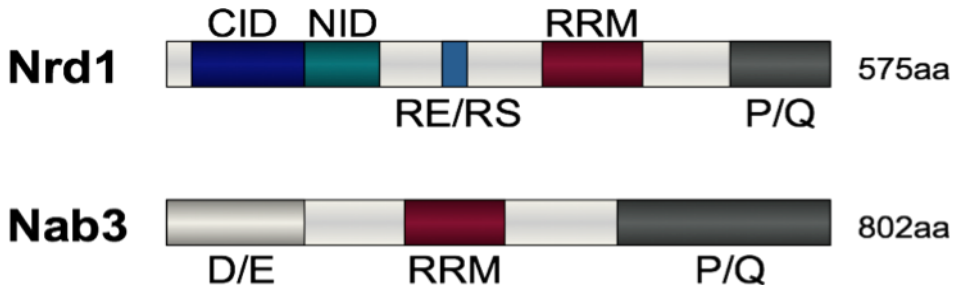

Figure 1.3 Domain organisation of Nrd1 and Nab3

CID - CTD interacting domain; NID - Nab3 interaction domain; RE/RS – region rich in arginine, glutamate and serine dipeptides; RRM - RNA recognition motif; P/Q -

proline and glutamine rich domain; D/E - stretch of acidic aspartic and glutamic residues.

Further studies have shown that Nrd1 interacts with the cap-binding complex (Cbp80) as well as the nuclear exosome (Rrp6) *in vivo,* linking the activities of Nrd1-Nab3 to nuclear surveillance (Vasiljeva and Buratowski, 2006). Genetic interactions indicated functionally significant interaction of Nrd1 and Nab3 with the RNA Pol II CTD. Point mutation in the CID or RRM of Nrd1 lead to synthetic growth defects in strains carrying only 10 out of the 26 heptad repeats that normally form the CTD (Conrad et al., 2000). Moreover, deletion of the CTD kinase (Ctk1) was able to suppress ts-lethal phenotypes of point mutations in the RRM of Nab3. Physical interactions between Nrd1 and the CTD were shown *in vivo* and *in vitro,* with a preference for the CTD phosphorylated at Ser-5 over Ser-2 phosphorylation (Vasiljeva et al., 2008a).

1.1.3.3 Other exosome cofactors - Rrp47 and Mpp6

Purification and mass spectrometry on epitope-tagged exosome components identified Lrp1/Rrp47 as an exosome-associated factor (Mitchell et al., 2003). Rrp47 is non-essential and present in approximately 20% of all exosome complexes; similar to Rrp6. A genetic interaction screen carried out with Rrp47 and the TRAMP component Air1 identified the non-essential, nuclear RNA binding protein Mpp6 as a further potential exosome cofactor (Milligan et al., 2008). Biochemical studies demonstrated a tight association of Mpp6 with the exosome and role for this cofactor in surveillance of pre-rRNA, pre-mRNA and cryptic intergenic transcripts could be proven.

The interaction of Rrp47 with the core exosome or Rrp6 is not RNA dependent and association of Rrp6 with the core exosome is not dependent on the presence of Rrp47. Analysis of RNA processing revealed however a clear requirement for Rrp47 in the processing of stable ncRNAs (see below; Mitchell et al., 2003). The human homologue of Rrp47 and the yeast protein were first characterised as factors

implicated in DNA double strand break repair (Erdemir et al., 2002; Yavuzer et al., 1998), however, the significance of this remains unclear (Houseley, 2008).

1.2 Activities of the nuclear RNA surveillance machinery

The exosome was first described as an exonuclease complex responsible for the correct processing of (pre-)rRNAs (Allmang et al., 1999a; Mitchell et al., 1997). Further studies revealed a plethora of functions for the exosome in the processing of stable ncRNAs such as rRNAs, snRNAs and snoRNAs, as well as in some special cases pre-mRNAs (Allmang et al., 1999a; Ciais et al., 2008; Roth et al., 2009; Torchet et al., 2002). Together with its various cofactors, the exosome is not only responsible for pre-RNA processing but also for surveillance and degradation of aberrant RNA species (reviewed by Houseley et al., 2006; Houseley and Tollervey, 2009 and illustrated in Figure 1.4).

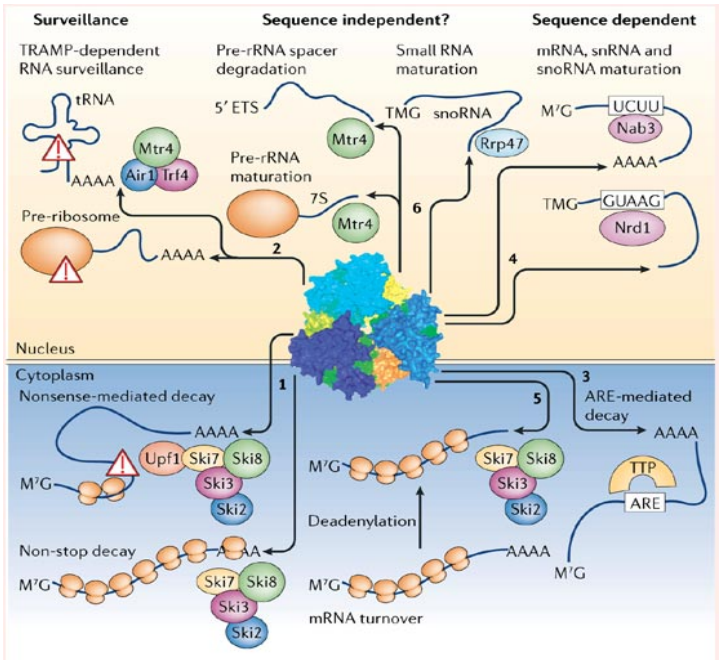

Figure 1.4 Activities of the exosome and its cofactors (Houseley et al., 2006)

1.2.1 Role of the exosome and cofactors in rRNA processing

Early studies on pre-rRNA processing characterised the function of the core exosome and the helicase Mtr4 in the 3' end formation of 5.8S rRNA (de la Cruz et al., 1998; Mitchell et al., 1996). Mutations in core exosome components or Mtr4 inhibit the conversion of 7S pre-rRNA to mature 5.8S. In addition to the core exosome, Rrp6 and Rrp47 are also involved in the maturation of 5.8S, but deletion of Rrp6 or Rrp47 lead to a different RNA processing phenotype, with accumulation of 5.8S which is extended by 30 nucleotides (5.8S+30). Final processing of the 6S pre-rRNA (5-8 nt extended 5.8S) seems to involve redundant activities, as it can be impaired by loss of any of several exonucleases: the exosome, Rex1, Rex2, Rex3 and the cytoplasmic factor Ngl2 (Allmang et al., 1999a; Thomson and Tollervey; van Hoof et al., 2000a). Furthermore, the exosome, Mtr4 and Rrp47 are responsible for the degradation of the 5' external transcribed spacer (5'ETS) fragment (Allmang et al., 1999a; de la Cruz et al., 1998; Mitchell et al., 2003).

1.2.2 Transcription termination and 3' end formation of sn(o)RNA genes

Processing of other stable ncRNAs such as snRNAs and snoRNAs has been investigated in the absence of the exosome and its cofactors (Allmang et al., 1999a; Grzechnik and Kufel, 2008; Mitchell et al., 2003; Steinmetz et al., 2001; van Hoof et al., 2000b). RNA Pol II transcribes snRNAs and snoRNAs, but their 3' ends are not formed by the conventional cleavage and polyadenylation machinery that is responsible for maturation of the 3' ends of mRNAs. Instead, it has been established that Nrd1 and Nab3 travel along with the transcribing polymerase mediated through the interaction between Nrd1 and the CTD. They initiate transcription termination once they encounter specific binding motifs within the RNA transcript and promote subsequent processing of the primary transcript by the exosome and TRAMP (Allmang et al., 1999a; Grzechnik and Kufel, 2008; Kim et al., 2006; Steinmetz and Brow, 2003; Steinmetz et al., 2001; Vasiljeva et al., 2008a).

The first indications for the existence of this pathway came from the observation that Nrd1, which can sequence specifically down regulate the expression of a reporter

mRNA, interacts with Sen1, which in turn is implicated in the processing of snoRNAs (Rasmussen and Culbertson, 1998; Steinmetz and Brow, 1998; Ursic et al., 1997). Microarray analysis revealed that the expression of ORFs that lie downstream of snoRNA genes is elevated in *nrd1* mutants, providing evidence for a role of Nrd1 in snoRNA transcription termination (Steinmetz et al., 2001). In the same study, transcription read-through phenotypes were described for several snoRNA genes in *nrd1* mutants. Mutants of Sen1, Nab3, Ctk1 or a truncation of the Pol II CTD exhibit similar snoRNA termination defects (Steinmetz et al., 2001). It had been proposed that snoRNA transcription termination could be facilitated by the cleavage and polyadenylation factor (CPF) through association with the snoRNP protein Nop1 (Morlando et al., 2004; Steinmetz and Brow, 2003). Several factors involved in mRNA 3‘ end formation can be detected over snoRNA as well as mRNA genes by chromatin immuno-precipitation (ChIP) (Kim et al., 2006). Convincing snoRNA transcription termination defects are, however, only visible in mutant strains of the Nrd1-Nab3-Sen1 complex (Kim et al., 2006). Thus, both ‘machineries’ for transcription termination of mRNA and sn(o)RNA genes are travelling together with the transcribing Pol II. The choice of which one is used on a particular gene is probably determined by signals residing within the RNA.

Downstream regions of snoRNA genes as well as spliceosomal U4 (snR14) contain similar sequence elements. These *cis* acting elements were found to be frequently altered in suppressor mutants in an ACT-CUP reporter assay that was used to study snoRNA transcription termination (Carroll et al., 2007; Carroll et al., 2004). Further investigations characterized the Nrd1 and Nab3 specific binding elements in depth, and band shift assays with recombinant proteins demonstrated that Nrd1 binds preferentially GUAA/G and Nab3 UCUU sequences *in vitro*.

Processing of the primary snRNA transcripts is dependent on endo- and exonucleases. The spliceosomal snRNAs contain a stemloop structure downstream of their mature 3’ end, which is a substrate for RNase III (Rnt1) cleavage (Chanfreau et al., 1997; Chanfreau et al., 1998a; Chanfreau et al., 1998b). For U1, U4 and U5 in particular it was shown that cleavage by Rnt1 generates an entry site for further

processing by the core exosome, Rrp47, Rrp6 and the exonuclease Rex2 (Allmang et al., 1999a; van Hoof et al., 2000a, b). Redundant functions of the exonucleases Rex1, Rex2 and Rex3 in the 3' end formation of U5 and other stable RNAs was demonstrated by Parker and co-workers (van Hoof et al., 2000a).

Final maturation of many snoRNA 3' ends is dependent on the nuclear exosome cofactors Rrp6 and Rrp47, as extended forms of box C/D snoRNAs and one HACA box snoRNA accumulated in the absence of either factor (Allmang et al., 1999a; Grzechnik and Kufel, 2008; Mitchell et al., 2003; van Hoof et al., 2000b). Longer 3' extended and polyadenylated species of single, intronic and polycystronic snoRNAs were also observed in mutants for the core exosome, demonstrating a general requirement for the nuclear exosome in 3' end formation of snoRNAs (Allmang et al., 1999a; van Hoof et al., 2000b). Further studies by Grzechnik and Kufel (2008) revealed that polyadenylation of the 3' extended pre-snoRNA by the TRAMP complex is part of the processing pathway. In the absence of Rrp6, various oligoadenylated and 3'extended pre-snoRNA species are stabilized and the precise oligoadenylation sites of the pre-snoRNAs were mapped (Grzechnik and Kufel, 2008).

1.2.3 Transcription termination, 3' end formation and regulated expression of some mRNAs is mediated by the nuclear RNA surveillance machinery

Even though the 3' ends of pre-mRNAs are formed by the cleavage and polyadenylation machinery (reviewed by Zhao et al., 1999), and mRNA turnover takes place in the cytoplasm, a few examples demonstrate a clear role for the nuclear RNA surveillance machinery in regulation of mRNA expression and 3' end formation (Arigo et al., 2006a; Ciais et al., 2008; Houalla et al., 2006; Reis and Campbell, 2007; Roth et al., 2009; Roth et al., 2005; Steinmetz et al., 2001; Torchet et al., 2002).

Strains carrying mutations in the cleavage and polyadenylation factors Rna14 and Rna15 are defective in pre-mRNA 3' cleavage, polyadenylation, and transcription termination (Minvielle-Sebastia et al., 1994). However, in these mutants functional mRNAs can be generated by activities of the nuclear exosome and Mtr4 (Torchet et al., 2002). Long extended read-through transcripts generated in these *rna14.1* and *rna15.2* strains are greatly stabilized by depletion of the core exosome (P_{GAL}*::RRP41*) or the TRAMP component Mtr4. Interestingly, the absence of Rrp6 from the *rna14.1* strain leads to a different phenotype, with stabilization of short polyadenylated pre-mRNAs that are functional for translation. Moreover, it was shown that the tight ts growth phenotype of both strains (*rrp6*Δ and *rna14.1*) was partially relieved in the double mutant, demonstrating that one major role of Rrp6 is in the degradation of defective pre-mRNAs. Northern blot analysis in this study demonstrated that the core exosome acts upstream of Rrp6 in the processing of pre-mRNAs in the absence of cleavage and polyadenylation, showing that the exosome generally possesses the ability to correctly process mRNA 3' ends.

The 5' UTR and the 5' coding region of the *NRD1* mRNA contain several Nrd1 and Nab3 consensus binding motifs and it has been shown that they are responsible for autoregulated expression of Nrd1 (Arigo et al., 2006a; Houalla et al., 2006; Steinmetz et al., 2001). Northern blot and microarray analyses show that *NRD1* mRNA levels are up-regulated in the absence of nuclear exosome components (*rrp41-1* and *rrp6*Δ) or exosome cofactors (*rrp47*Δ and *trf4*Δ), in mutants of the cleavage and polyadenylation machinery (*ysh1-12*) and in mutants of Nab3, Sen1 or Nrd1 (Arigo et al., 2006a; Garas et al., 2008; Houalla et al., 2006; Steinmetz et al., 2001). *NRD1* expression is regulated by Nrd1-Nab3-directed premature transcription termination, with the possible involvement of the 3' end processing endonuclease Ysh1. Transcription termination is guided by the consensus Nrd1 and Nab3 binding motifs within the *NRD1* mRNA, as mutations in these lead to mRNA accumulation to a similar extent as observed in the *nrd1* or *nab3* mutants (Arigo et al., 2006a). Consistent with the premature transcription termination model, truncated Nrd1 mRNA 5' fragments, heterogeneous in size, are detectable in strains carrying *rrp6*Δ*, rrp47*Δ *or rrp41-1* mutations, whereas *rat1-1 xrn1*Δ double mutants stabilize the

downstream cleavage product of the mRNA (Garas et al., 2008; Houalla et al., 2006). Furthermore, ChIP and transcription run-on experiments showed that mutants of Nrd1 and Nab3 exhibit a relative increase in Pol II occupancy over the 3' end of the mRNA, compared to the WT which generally shows a low Pol II occupancy (Arigo et al., 2006a).

Cth2 is an ARE-binding protein functioning in mRNA degradation in the cytoplasm. The *CTH2* mRNA was found to be over-expressed in microarray analyses of strains lacking the nuclear exosome component Rrp6 (Houalla et al., 2006). Studies by Ciais *et al.* (2008) showed that *CTH2* mRNA is not generated by the conventional cleavage and polyadenylation machinery, but is processed from a 3'-extended primary transcript by the exosome, TRAMP and Nrd1/Nab3/Sen1 complexes (Ciais et al., 2008). Northern analyses revealed that exosome mutants accumulate a 3'-extended *CTH2* mRNA transcript, which contains numerous Nrd1 and Nab3 binding sites. Consistent with this model, *CTH2* mRNA levels were not altered by mutations in 3'-cleavage and polyadenylation factors, Rna14, Rna15 and Pap1. Moreover, mutations in the TRAMP or the Nrd1/Nab3/Sen1 complexes also lead to a vast accumulation of the 3'-extended *CTH2* pre-mRNA. The polyadenylation site was identified as lying in a $[GU_{(3-5)}]_{(5)}$ repeat, but the significance of this motif is unclear.

1.2.4 Pervasive transcription: Stabilisation, degradation and functions of cryptic ncRNAs

1.2.4.1 Identification of cryptic ncRNAs

Recently it has become apparent that most of the genome is actively transcribed by RNA Pol II giving rise to cryptic transcripts. The resulting RNAs are, however, not readily detectable in WT cells, as they are efficiently degraded by the nuclear surveillance machinery (reviewed by Berretta and Morillon, 2009; Houseley and Tollervey, 2008; Jacquier, 2009).

The first indications that the transcriptome exceeds known mRNAs and stable ncRNA such as rRNA, sn(o)RNAs came from SAGE (serial analysis of gene expression) and whole genome microarray analysis carried out in strains deleted for the nuclear exosome component Rrp6 (Davis and Ares, 2006; Wyers et al., 2005). These approaches revealed that transcripts are emerging from numerous intergenic regions and that these RNAs are stabilised in the absence of the nuclear exosome. Further analysis revealed that these relatively short transcripts (200-800 nts) are capped, heterogeneous in length due to different 3' ends, polyadenylated and also accumulate in the absence of the poly(A) polymerase Trf4. Due to their appearance only in mutants of the surveillance machinery they were termed cryptic unstable transcripts (CUTs) (Wyers et al., 2005). In the study carried out by Davis and Ares these transcripts were found to be clustered around promoter regions of annotated genes and were therefore named short promoter associated transcripts (Davis and Ares, 2006). As expected, they were especially enriched in loci between divergent genes. From the experimental design it is not clear, however, if the small ncRNAs are expressed in sense or antisense direction with respect to the mRNA. Northern hybridisation in this study was used to detect the promoter associated RNAs by random primed probes, which are not strand specific (Davis and Ares, 2006).

Regions depleted of nucleosomes generally allow the transcription machinery to initiate and generate transcripts (see Fig. 1.5) (reviewed by Berretta and Morillon, 2009). This is the case for the nucleosome-depleted regions (NDR) around mRNA transcription start sites. However, if nucleosomes are not correctly repositioned after transcription or in the absence of the chromatin remodeler Isw2, cryptic transcription also can initiate (Whitehouse et al., 2007). Aberrant transcription within coding regions has also been observed in mutants of transcription elongation factors Spt6 and Spt16 (Cheung et al., 2008; Kaplan et al., 2003). Winston and co-workers showed that some of these cryptic transcripts contain ORFs and give rise to polypeptides suggesting that under altered genetic or physiological conditions alternative genetic information might be expressed.

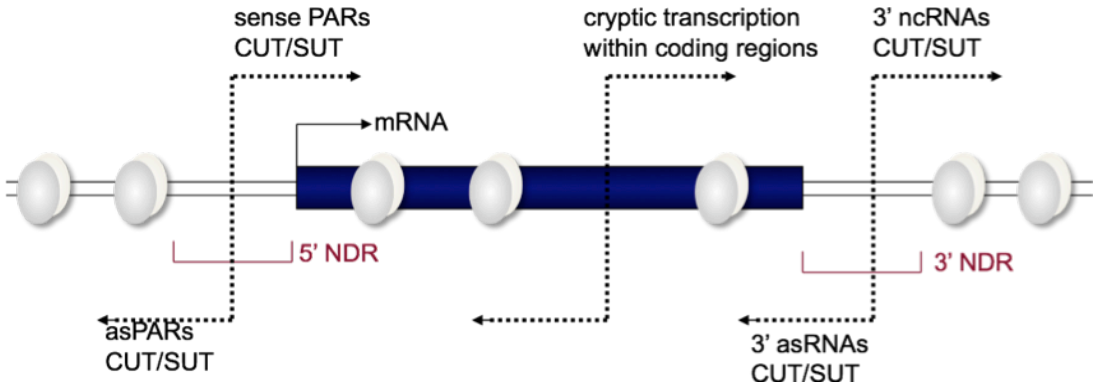

Figure 1.5 Image adapted from (Berretta and Morrillon, 2009)

More recently, two studies have generated high-resolution profiles of the cryptic transcriptome in yeast (Neil et al., 2009; Xu et al., 2009). One study employed tiling arrays as well as high-throughput sequencing by 3' long SAGE approaches of a nuclear RNA fraction that was highly enriched for CUTs from a *trf4*Δ *rrp6*Δ double mutant strain (Neil et al., 2009). This method identified 1496 CUT cluster that did not correspond to any annotated feature. The second approach conducted sensitive tiling arrays and compared the transcriptome of an *rrp6*Δ mutant and of WT yeast under different nutrient conditions (Xu et al., 2009). In this study 7272 transcripts were identified, 71% of which were accounted for by ORFs and stable ncRNAs (rRNA, tRNA, sn(o)RNA) and 13% by CUTs. In addition, a new class of ncRNAs was defined as SUTs (stable unannotated transcripts); these do not carry features of an ORF but are expressed at similar level in WT and *rrp6*Δ cells and account for 12% of all transcripts. Comparison of the CUTs identified by both approaches shows only partial overlap, presumably due to the different techniques and mutant strains utilized. However, in both studies it was observed that the pervasive transcription is not random, but clustered at NDRs, preferably at the 5' end of genes (68%) but also at the 3' end (32%; Lee et al., 2007; Xu et al., 2009). Remarkably, CUTs predominantly (95%) orient antisense to the ORF, with only 5% in sense orientation to an mRNA generating overlapping transcripts (Neil et al., 2009; Xu et al., 2009). One major conclusion from these studies was that promoters frequently function bidirectionally. Even though the identified CUTs do not perfectly overlap between both studies, the bidirectional character of most of the promoters that generate

promoter associated RNAs and the large number of initiation events from shared NDRs are consistent findings.

1.2.4.2 Transcription termination and degradation of cryptic ncRNAs

Following the initial characterisation of pervasive transcription, it was shown that Nrd1 and Nab3, required for transcription termination and subsequent processing of sn(o)RNAs by the TRAMP exosome pathway, are also involved transcription termination of CUTs and promoter associated RNAs (Arigo et al., 2006b; Thiebaut et al., 2006).

Libri and colleagues further characterised a CUT that they identified in their initial SAGE analysis, *NEL025c* (Thiebaut et al., 2006). This CUT is expressed as a set of heterogeneous short transcripts (NEL025s) and a distinct longer transcript (NEL025L). Both forms are stabilised by either *rrp6*Δ or *trf4*Δ, but only the short form is polyadenylated by Trf4, whereas the longer transcripts seem to obtain a poly(A) tail from conventional Pap1 (Thiebaut et al., 2006; Wyers et al., 2005). Attempts to 'stabilize' the NEL025 CUT by addition of artificial *CYC1* promoter and terminator elements failed, leading to the conclusion that sequence elements within the transcripts must confer instability. A large number of Nrd1 and Nab3 binding motifs were found in *NEL025* and, indeed, the short form NEL025s was susceptible to Nrd1-Nab3 dependent termination. Depletion of either Nrd1 or Nab3 caused the exclusive accumulation of the longer, read-through NEL025L form and in a double mutant strain (P_{GAL}*::NRD1 rrp6*Δ) hardly any NEL025s could be detected. This is a clear indication that NEL025L is not a precursor to NEL025s, but a read-through transcript of the shorter RNA.

Similar observations were made for RNAs running antisense to known ORFs, which were identified by microarrays in a temperature sensitive *nab3-11* mutant (Arigo et al., 2006b). Various ncRNAs originating within or at the 3' end of ORFs were identified that run antisense to the mRNA (further referred to as asRNAs). These transcripts were short, polyadenylated, heterogeneous in size and readily detectable

in *rrp6, trf4* as well as *rrp47* mutants. Interestingly, temperature sensitive strains bearing point mutations in the RRM of Nrd1 or Nab3 (*nrd1-102* and *nab3-11*) exhibited a different phenotype, as they accumulated longer read-through products of a distinct length. In addition, a synthetic growth defect was observed between *nrd1* and *rrp47* mutants and northern hybridisation revealed that the double mutant exhibits the same transcription read-through defect as the *nrd1* mutant alone. Therefore these experiments together with the analysis of the P_{GAL}*::NRD1 rrp6*Δ double mutant (Thiebaut et al., 2006) established that Nrd1 and Nab3 act in transcription termination of short asRNAs and CUTs, directing them to the TRAMP/exosome degradation pathway.

1.2.4.3 Functions of cryptic intergenic and antisense ncRNAs

Several intergenic or antisense ncRNAs have been functionally analysed in yeast, proving that these transcripts do not represent transcriptional junk but participate in the regulated expression of other genes (reviewed by Berretta and Morillon, 2009; Harrison et al., 2009). Cryptic RNA expression affects transcription of other genes by various mechanisms.

Transcriptional interference mechanisms have been described in which transcription of the cryptic transcripts covers the promoter region of a downstream gene thereby preventing expression of that gene (several examples in Berretta and Morillon, 2009; Harrison et al., 2009; Jacquier, 2009). *SRG1* ncRNA is transcribed under high serine concentrations through actions of the serine dependent transcription activator Cha4 preventing transcription of the downstream mRNA gene *SER3* encoding for a factor involved in serine biosynthesis (Martens et al., 2004; Martens et al., 2005). This elegant feedback loop was the first example of an ncRNA directly regulating expression of an mRNA.

Transcription through a DNA region inevitably leads to chromatin remodelling events as nucleosomes need to be evicted in order for the polymerase to pass and repositioned afterwards (reviewed in Cairns, 2009). Actively transcribed regions of

the genome contain different modification on the histone tails than silent regions and it is believed that these modifications help determine the accessibility of the DNA and thereby regulate gene expression. In fission yeast *S. pombe* the RNAi machinery and siRNAs are involved in heterchromatic silencing (reviewed by Grewal, 2010). *S. cerevisiae* lacks the RNAi system but there is mounting evidence that ncRNAs take part in regulating chromatin structure (reviewed by Houseley and Tollervey, 2008). Genetic interactions between the surveillance machinery and chromatin remodelling factors suggested an active role for these proteins in chromatin structure rearrangements. Expression of several cryptic ncRNAs has been shown to alter the chromatin state of the locus. In some cases this leads to repression of the cognate mRNAs (several examples in Harrison et al., 2009) or conversely opening of the chromatin structure of the region promoting expression of other genes (Uhler et al., 2007).

Chromatin within the spacer region of the rDNA repeats is transcriptionally silenced by the action of the histone deacetylase Sir2. In the absence of Sir2 or in mutants of the surveillance machinery (*trf4*Δ*, rrp6*Δ*,* P_{GAL}*::NRD1*) several cryptic Pol II transcripts were detectable (Houseley et al., 2007b; Vasiljeva et al., 2008b). The identified transcripts were polyadenylated and showed heterogeneous 3' ends (Houseley et al., 2007b). Buratowski and co-workers showed that transcription termination of these cryptic transcripts is mediated through the Nrd1 termination pathway. Furthermore they demonstrated by insertion of a reporter into the silenced region, that upon inactivation of the surveillance machinery, meaning expression of the ncRNAs, silencing was relieved (Vasiljeva et al., 2008b). Interestingly, Nrd1 but not Trf4 mutants exhibited increased recombination between rDNA repeats as shown by colony sectoring assays with a reporter (Houseley et al., 2007b; Vasiljeva et al., 2008b). Moreover, Trf4 was proven to function in rDNA copy number control (Houseley et al., 2007b). These findings link ncRNA surveillance activities to chromatin structure and integrity.

In a different genomic location, Trf4 contributes to the repression of mRNAs through chromatin modifications induced by an ncRNA (Houseley et al., 2008). The genes of

the *GAL* cluster (*GAL1, GAL7* and *GAL10*) are repressed in glucose media but induced in galactose containing media through binding of the transcription factor Gal4. Surprisingly the 3' region of the *GAL10* gene showed chromatin marks of active transcription (H3K4me3) in the repressive state. Vogelauer and co-workers demonstrated that an ncRNA is transcribed in the *GAL* cluster from the 3' end of *GAL10* when mRNA transcription is repressed (Houseley et al., 2008). ncRNA transcription was shown to recruit the histone methyltransferase Set2 and change the chromatin state in the region thereby enhancing mRNA repression in low levels of glucose.

1.3 RNA Pol III genes: Transcription, processing and surveillance

1.3.1 RNA Pol III transcription

In eukaryotes RNA Pol III transcribes a large group of stable ncRNA genes, such as 5S rRNA, all tRNAs, RNase P RNA (*RPR1*), U6 snRNA, *SCR1* (or 7SL, the RNA component of the signal recognition particle) and recently a few other RNAs (snR52, RNA170, ZOD1) were identified as Pol III transcripts in *S. cerevisiae*. Unlike all other organisms tested, the RNA component of RNase MRP (*NME1*) is not a Pol III but a Pol II transcript in *S. cerevisiae* (reviewed by Dieci et al., 2007).

RNA Pol III consists of 17 individual factors in yeast, some of which are specific to Pol III while others are related to subunits of Pol I and Pol II (reviewed by Geiduschek and Kassavetis, 2001). Transcription initiation by Pol III is dependent on the transcription factors TFIIIC and TFIIIB (reviewed by Dieci et al., 2007; Geiduschek and Kassavetis, 2001). Specific promoter elements (boxA and boxB) reside within the coding region of the gene and are recognised by the multi-subunit factor TFIIIC. Following TFIIIC binding, TFIIIB is recruited to a TATA-like element situated upstream of the transcription start site. Promoter elements were first characterised for tRNAs but these are not strictly conserved to other Pol III genes. U6, for instance, possesses a downstream boxB element whereas for the *RPR1* gene, both boxA and boxB are located within the 5' leader sequence. Promoter elements of

the 5S gene are entirely different, as they consist of an internal boxA and an intermediate element, called boxC, but not a boxB. Recognition of the intermediate element requires a specific recognition factor, TFIIIA, which is further recognised by TFIIIC. Given these differences, the promoters are classified as Type 1 (5S) or Type 2 (tRNAs, SCR1, RPR1). In humans, further upstream enhancer elements have been characterised, they make up Type 3 promoters (Figure 1.6).

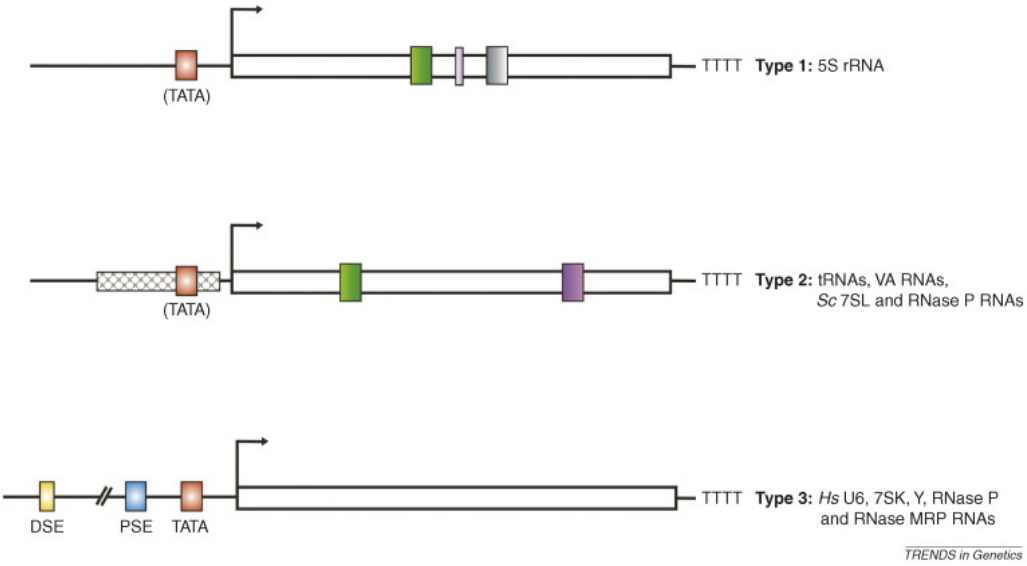

Figure 1.6 Pol III promoter elements (Dieci et al., 2007)

Transcription termination of RNA Pol III is unique, as it only requires a short run of T residues (T≥4 in vertebrates and T≥5 in yeast) and no accessory factors. The base pairing interaction between the oligo U stretch of the nascent transcript and the encoding A stretch on the DNA is so weak that the polymerase just slips off and thereby terminates (reviewed by Geiduschek and Kassavetis, 2001).

1.3.2 Processing of Pol III transcripts

1.3.2.1 RNA modifications

Following transcription in or around the nucleolus, RNA Pol III transcripts undergo extensive processing and maturation (reviewed by Phizicky and Hopper, 2010; Walker and Engelke, 2006). Extension at the 3' and 5' ends need to be removed and numerous base modifications are introduced, which have been extensively studied

for tRNAs (Czerwoniec et al., 2009). A total of 92 different base modifications have been found to date in various species and are listed in the RNA modification database (http://biochem.ncsu.edu/RNAmods). For *S. cerevisiae* 25 modifications were identified for 34 different tRNA species (Phizicky and Hopper, 2010). These include base methylation, acetylation, pseudo-uridylation and many more, which influence translation at multiple levels by enhancing stability or promoting export (reviewed in Phizicky and Hopper, 2010). Many modification reactions take place in the nucleus but are distributed between various subnuclear locations, such as the nucleolus, the nucleoplasm or the inner nuclear membrane. The reasons for the subnuclear distribution of tRNA modification activities are unclear at present. Some modification reactions can only occur after tRNAs have been spliced, which takes place in the cytoplasm (see below), and are therefore exclusively cytoplasmic (reviewed in Phizicky and Hopper, 2010).

Despite the influence that modifications can have on tRNA stability and their evolutionary conservation, very few are essential for cell viability. Notably, loss of modifications in the region around the tRNA anticodon causes severe growth defects and lack of the factors establishing these modifications is lethal reflecting the importance of translational fidelity. For instance yeast strains lacking I_{34}, from conversion of adenosine to inosine by deamination, are inviable (Gerber and Keller, 1999). Inosine (read as G in a sequencing reaction) at position 34 allows non-Watson-Crick base pairing and is called the wobble position nucleotide (Crick, 1966). Wobble base pairs are I-A, I-C and I-U; therefore this modification allows recognition of more than one mRNA codon by the modified tRNA.

1.3.2.2 RNase P

Ribonuclease P (RNase P) is an RNP endoribonuclease, responsible for removal of the 5' leader sequences of tRNAs. It is conserved through evolution and in yeast the RNA moiety (*RPR1*) and nine protein subunits (Pop1, Pop3, Pop4, Pop5, Pop6, Pop7, Pop8, Rpr2 and Rpp1) have been shown to be essential for viability (reviewed by (Walker and Engelke, 2006). The bacterial RNase P RNA (M1 RNA) was shown

to catalyse pre-tRNA cleavage alone *in vitro* under high Mg^{2+} ion concentrations and was therefore the first example of an RNA enzyme or ribozyme (Guerrier-Takada et al., 1983). *In vivo* M1 RNA is accompanied by a small protein subunit (C5), addition of which led to several fold increase of catalytic activity in the *in vitro* assay. In contrast, eukaryotic RNase P RNA has never been shown to be catalytically active on its own (Lee and Engelke, 1989).

Yeast *RPR1* RNA is transcribed as a precursor transcript containing a 5' leader and 3' trailer sequence, which is assembled into the RNP prior to processing (Srisawat et al., 2002). Engelke and colleagues have mapped the 5' end by primer extension and the 3' end by S1 nuclease digestion. A single 5' extended form (by 84 nts) and several 3' extended spices (16-30 nts) were identified (Lee et al., 1991). These observations led them to conclude that the mature 3' end was generated in a stepwise process, whereas 5' leader removal was performed by a single cleavage event; the enzymatic activity responsible for this however has not been identified to date. Regarding formation of the 3' end, the Parker lab had shown that three RNase D type exonucleases (Rex1, Rex2 and Rex3) function redundantly in part of the processing, as the triple deletion mutant accumulates *RPR1* with a small 3' extension (van Hoof et al., 2000a). Redundancy seems to be a general feature of 3' end processing events of stable RNAs and it is conceivable that other exonucleases function redundantly with the Rex proteins in the removal of the 3' trailer of pre-*RPR1*. Pre-*RPR1* has been shown to be catalytically active in the processing of pre-tRNAs *in vitro* (Srisawat et al., 2002). Use of an RNA affinity tag allowed purification pre-*RPR1* and showed that this active holoenzyme contains all subunits apart from Pop3 and Rpr2.

1.3.2.3 tRNA processing

tRNA biogenesis is very well studied in several organisms. The primary tRNA transcript contains a 5' leader sequence, a 3' trailer and some species also have intronic sequences that need to be removed by splicing. In most organisms (apart from some bacteria and archea) tRNA genes do not encode terminal CCA

nucleotides. Following processing and modification, the CCA is added before the tRNA can be charged with an amino acid (reviewed in Phizicky and Hopper, 2010). These processing steps have been shown to occur in different cellular compartments and in a strict order, probably representing quality control measurements. A schematic representation of the tRNA maturation pathway is displayed in Figure 1.7.

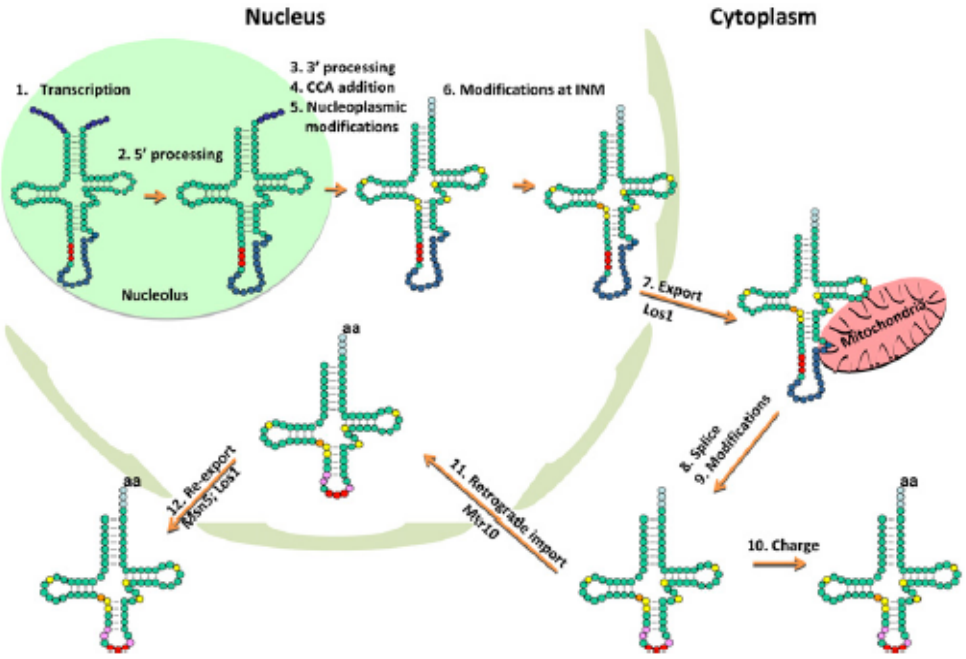

Figure 1.7 tRNA processing pathway (Phizicky and Hopper, 2010)

In yeast, the first processing step after transcription is the removal of the 5' leader sequence by the endonuclease RNase P, which takes place in the nucleolus (see above; Bertrand et al., 1998). Processing of the 5' end is followed by removal of the 3' trailer sequences; only for $tRNA^{Trp}$ it was shown that 3' processing precedes 5' processing (Kufel and Tollervey, 2003). 3' end formation of tRNAs is mediated by endonucleolytic cleave and subsequent exonuclease trimming in both bacteria and eukaryotes (reviewed in Phizicky and Hopper, 2010). In yeast and humans, a metallo beta-lactamase family protein, tRNase Z (Trz1 in yeast, ELAC2 in humans), encodes the endonuclease that cleaves the tRNA 3' trailer. Nashimoto and co-workers have expressed Trz1 and ELAC2 in *E.coli* and shown that the purified recombinant proteins can process tRNAs *in vitro* (Takaku et al., 2003). *In vivo,* however,

exonucleases also contribute to the processing of the 3' trailer. Many contributions to the understanding of tRNA 3' end formation in yeast have been made by Wolin and co-workers, who showed that the La protein (Lhp1 in yeast) promotes 3' cleavage by an endonuclease presumably by covering the 3' trailer and guiding the endonuclease to the correct cleavage site (Yoo and Wolin, 1997). In this study, *in vitro* tRNA transcription/processing assays were carried out in *lhp1*Δ cell extracts, revealing that in the absence of La exonuclease trimming generates mature tRNA 3' ends. Following this initial work, they showed that Rex1 is the major exonuclease responsible for trailer trimming (Copela et al., 2008). Furthermore Rex1 (*rna82*) was shown to be responsible for the 3' trimming of the first part of the dicistronic $tRNA^{Arg3}$ (Piper and Stråby, 1989; van Hoof et al., 2000a). Localization studies on Trz1 and Lhp1 suggest that these processes are probably nucleoplasmic (Huh et al., 2003).

Following correct 3' end maturation of the tRNAs the CCA nucleotidyl transferase or CCA adding enzyme adds two C and one A residue to the pre-tRNA in the nucleus (reviewed by Martin and Keller, 2007; Phizicky and Hopper, 2010). In vertebrate cells this processing step is required for efficient tRNA export. There is also evidence for yeast CCA addition to be nuclear, as end-matured and intron containing tRNAs in the nucleus generally contain a CCA. In addition to the nuclear CCA adding enzymes, cytoplasmic forms have been found. It is suggested that the cytoplasmic pool serves as repair enzymes for otherwise mature cytoplasmic tRNAs.

In all sequenced archaea and eukaryotes some tRNA species contain intervening sequences that need to be removed post transcriptionally. These are generally located downstream of the anticodon loop between nucleotides 37 and 38. Splicing of tRNA introns is a lot simpler than the sophisticated actions of the mRNA spliceosome (Figure 1.8). The splicing endonuclease, consisting of four factors (Sen2, Sen15, Sen34 and Sen54 in yeast) cuts out the intron from the tRNA. Sen2 and Sen34 are related and have been shown to be the catalytic subunits of the complex (Li et al., 1998). Excision of the intron leaves a 2'-3' cyclic phosphate on the 5' fragment and a 5'-OH group on the 3' fragment. Joining of the two ends requires several enzymatic

activities, hydrolysing the 2'-3' cyclic phosphate to generate a 2' phosphate, phosphorylation of the 3' fragment by an RNA kinase and finally ligation of the two halves (Greer, 1986). In yeast one protein, Trl1, is responsible for these reactions; in vertebrates the factor(s) responsible are not known (Phizicky et al., 1986).

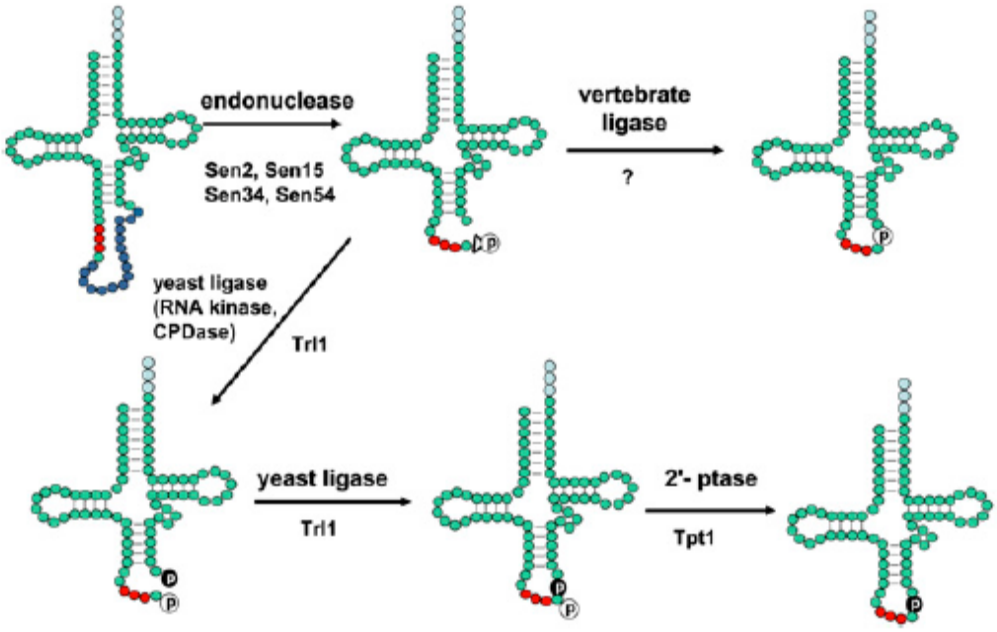

Figure 1. 8 tRNA splicing (Phizicky and Hopper, 2010)

Major differences have been found in various organisms concerning the location of tRNA splicing. In vertebrates this process is clearly nuclear while in yeast it takes place at the outer mitochondrial membrane in the cytoplasm. For plants opposing results have been published and it has not been resolved whether tRNA splicing in plants is nuclear or cytoplasmic (reviewed by Phizicky and Hopper, 2010).

The first evidence that tRNA splicing in yeast might be cytoplasmic came from genome wide protein localization studies, since the splicing endonuclease complex could only be detected at the outer mitochondrial membrane (Huh et al., 2003). Endo and co-workers could furthermore demonstrate that unspliced pre-tRNAs accumulating in the cytoplasm are indeed processing intermediates (Yoshihisa et al., 2007; Yoshihisa et al., 2003). The export factor Los1 is responsible for transport of the unspliced und uncharged tRNAs to the cytoplasm. Following splicing they are aminoacylated and can participate in translation.

Given that tRNA processing takes place at various locations in the cell, a transport system is needed to account for this mobility. Transportation of tRNA is not unidirectional from the nucleus to the cytoplasm (as reviewed in Phizicky and Hopper, 2010). The retrograde tRNA transport system is able to react to nutrient availability and import tRNAs via the ß-importin factor Mtr10. Re-export of any charged and properly matured tRNA can be carried out by the export factors Msn5 and Los1.

1.3.2.4 5S rRNA processing

5S rDNA genes are located within the each of rDNA repeats on chromosome XII in yeast. The 5S rRNA is transcribed from the opposite strand to the 35S transcription unit that encodes the other rRNAs, which are transcribed by RNA Pol I. The primary 5S transcript contains a 3' oligo U tract and is 3' matured, probably by redundantly acting exonucleases. To date only *rex1*Δ mutants have exhibited a 5S processing defect. Parker and colleagues have shown by northern analyses and RNase H digests, that in the absence of Rex1 5S extended by 3 nt at the 3' end accumulates at steady state, and no mature 5S is formed (van Hoof et al., 2000a). Previous studies had shown by pulse chase analyses in a Rex1 mutant allele (*rna82.1*) that pre-5S with extensions as long as 13 nts accumulate (Piper et al., 1983).

1.3.3 Surveillance of defective Pol III transcripts

Surveillance of transcripts generated by RNA Pol III has not been as extensively studied as for Pol II derived RNAs. However, a few substrates haven been examined in detail. Anderson and co-workers screened for mutants that could suppress the growth defect caused by a *trm6* mutation attributable to the lack of m^1A_{58} and the resulting instability of $tRNA_i^{Met}$ (Kadaba et al., 2004). Suppressors identified in this screen, Rrp44 and Trf4, were involved in degradation of the defective RNA, as mutations in both factors restored normal levels of the mutant tRNA in the *trm6* mutant strain and relieved the growth defect. Following these initial studies, they showed by sequencing that Trf4 adds A-tails to hypomodified pre-$tRNA_i^{Met}$ *in vivo*

(Kadaba et al., 2006). Mutation of a conserved polymerase core motif DXD to AXA (*trf4 DADA*) abolished polyadenylation of the defective $tRNA_i^{Met}$ (Kadaba et al., 2006). Detailed *in vitro* studies showed that reconstituted as well as TRAMP4 complexes purified from yeast selectively polyadenylated *in vitro* transcribed, unmodified $tRNA_i^{Met}$ compared to native yeast $tRNA_i^{Met}$ (Vanacova et al., 2005). This polyadenylation activity was lost in the *trf4 DADA* mutant. Furthermore they could show that polyadenylation of unmodified $tRNA_i^{Met}$ by TRAMP stimulated degradation by Rrp6 containing exosomes, demonstrating that poly(A) tails serve as a signal for exosome degradation. In addition, the exonuclease activity of Rrp44 was shown to be responsible for degradation of pre-$tRNA^{Ser}$ *in vitro* (Schneider et al., 2007). Rrp44 was also able to distinguish between hypomethylated and fully modified $tRNA_i^{Met}$ independently from TRAMP and to partially degrade the RNA. Similarly to the observations made for Rrp6 exosomes, addition of TRAMP stimulated Rrp44 degradation activity and lead to complete hydrolysis of the tRNA substrate (Schneider et al., 2007). In addition, intron-containing pre-tRNAs accumulate in the absence of Trf4, indicating that these species are usually targeted for degradation by TRAMP4 (Copela et al., 2008).

Mature tRNAs as well as unmodified pre-tRNAs have been identified as targets for surveillance by the rapid tRNA decay pathway (RTD). This, however, degrades tRNAs from the 5' end, mediated by the exonucleases Rat1 and Xrn1 (Alexandrov et al., 2006; Chernyakov et al., 2008).

Other Pol III derived RNAs have been identified as targets for the nuclear RNA surveillance machinery. Libri and colleagues demonstrated that in the absence of the nuclear exosome component Rrp6 polyadenylated forms of U6 snRNA and 5S rRNA accumulate. Heat inactivation of a temperature sensitive mutant of Pap1 (poly(A) polymerase of the cleavage and polyadenylation machinery) did not abolish polyadenylation of these spices. Therefore they concluded that a different entity, Trf4, must be responsible for this (Wyers et al., 2005).

Additional studies identified a 3' truncated, aberrant form of 5S (5S*), which presumably results from partial 3' end degradation by exonucleases (Kadaba et al., 2006). Truncated 5S* accumulated in mutants of the TRAMP and exosome complexes (*rrp6*Δ*, trf4*Δ and *rrp44-20*) and was polyadenylated in these mutants, except for *trf4*Δ, demonstrating that the exosome and Rrp6 are responsible for its degradation. In the same study, the role of these factors was investigated concerning degradation of an unstable U6 mutant lacking 14 nts (*U6* Δ*59-72*). Strains expressing plasmid borne *U6* Δ*59-72,* in addition to the genomic *SNR6* gene, accumulated the mutant U6 in the absence of the surveillance machinery (*rrp6*Δ and *trf4*Δ) or in the polyadenylation deficient *trf4 DADA* strain.

1.4 Crosslinking approaches

Many different proteins participate in the maturation of RNAs and the associated quality control mechanisms. Identification of potential targets for RNA binding proteins, especially for mRNAs, has generally used microarray techniques to display relative RNA abundances under defined growth condition or mutant backgrounds. Studies of direct interactions between RNAs and proteins were generally limited to individual RNA species. Recent crosslinking approaches have brought huge advances in the identification of targets for RNA binding proteins genome wide. Darnell and colleagues developed assays based on UV crosslinking and immunoprecipitation CLIP (cross linking and immunoprecipitation), which can be used to identify direct targets for RNA binding proteins (Ule et al., 2003). UV crosslinking of proteins to RNA was performed on brain tissue, the protein of interest was immunoprecipitated with an antibody and covalently associated RNAs were cloned and sequenced. The method gained popularity because it was applicable to other tissues, and improvements have been made, for instance the use of next generation sequencing technologies to identify targets genome wide (Granneman et al., 2009; Ule et al., 2005).

While CLIP would be applicable to yeast, specific antibodies would be needed for each protein. The advantages of yeast as a model organism include the possibility of

rapidly generating mutants and epitope-tagged proteins. Dr. Sander Granneman in the lab made use of this and designed the HTP tag (His-TEV-ProtA), which is closely related to the TAP tag widely used in yeast biochemistry for tandem affinity purifications. UV crosslinking and immunoprecipitation experiments were adapted to the tagged proteins and the method was named CRAC (Crosslinking and analysis of cDNAs; (Granneman et al., 2009). The protocol allowed tandem affinity purification of the crosslinked RNA protein complexes under denaturing conditions. The method is further introduced in Chapter 3, while Chapter 2 contains the detailed protocol of the CRAC approach.

1.5 Telomerase

The linear ends of eukaryotic chromosomes are protected by a cap structure consisting of a repetitive DNA sequence that is covered by proteins, the telomere (from Greek telos=end). Therefore they can be distinguished from double strand DNA breaks that require repair. Telomerase is the ribonucleoprotein particle responsible for maintenance of the telomere by elongating the single stranded 3' end of the DNA. A reverse transcriptase (Est2 in yeast) uses the RNA moiety (*TLC1* in yeast) as a template to add the telomere repeat sequence (TTGGGG in yeast). This extended single strand provides a template for DNA polymerase, thereby preventing loss of genetic information from the ends of the chromosomes during each round of DNA replication.

In vitro, only *TLC1* and Est2 are required for telomerase activity (Counter et al., 1997) but three additional factors Est1, Est3 and Cdc13 are required for telomere maintenance *in vivo*. Mutations in any of the above factors leads to progressive telomere shortening, a phenotype termed 'ever shorter telomeres' and eventually cell death. Telomerase RNA serves as the template for DNA repeat addition as well as a scaffold for the other telomerase factors (Zappulla and Cech, 2004). The catalytic subunits are highly conserved, whereas telomerase RNAs vary greatly in sequence and length. Phylogenetic analyses and structure probing have been carried out to

determine the structure of budding yeast *TLC1* RNA (Dandjinou et al., 2004; Zappulla and Cech, 2004).

In budding yeast, *TLC1* is transcribed by RNA Pol II and two major forms of the RNA have been identified *in vivo* (Chapon et al., 1997). Mature *TLC1* RNA is approximately 1.2 kb long, carrying a 2,2,7-trimethyl guanosine cap at the 5' end and an Sm-protein binding site adjacent to the 3' end (Seto et al., 1999). The mature 3' end of *TLC1* RNA had been mapped on RNAs purified with active telomerase (Bosoy et al., 2003). In addition, a polyadenylated longer form of *TLC1* has been identified, which was hypothesised to be the precursor to *TLC1* (Chapon et al., 1997). Cech and co-workers identified this polyadenylated species in the 'unsuccessful' attempt to map the 3' end by polyadenylation of total RNA followed by oligo(dT) primed RT PCR. The control, which was not treated with poly(A) polymerase, gave the same *TLC1* species identified as in the polyadenylated fraction. Therefore, they concluded that this longer pre-*TLC1* species is genuinely polyadenylated. Furthermore they showed that upon heat inactivation of a temperature sensitive allele of poly(A) polymerase (*pap1-5*) the signal for this RNA was lost. However, it was unclear which processing events convert the polyadenylated pre-*TLC1* to mature *TLC1* or whether pre-*TLC1* is indeed the precursor to mature *TLC1*.

In the distantly related yeast *S. pombe*, where telomerase RNA (*TER1*) is similar in size to *S. cerevisiae TLC1,* a larger polyadenylated species has also been identified (Leonardi et al., 2008; Webb and Zakian, 2008). Recently, Baumann and colleagues dissected the mechanism by which the *TER1* 3' end is formed in *S. pombe* (Box et al., 2008). They noted that 5% of poly(A) RNAs recovered lacked part of the sequence adjacent to the Sm binding site at the 3' end. This region contains matches to the 5' splice site, 3' splice site and branch point consensus sequences. Indeed, maturation of *TER1* was dependent on spliceosomal cleavage at the 5' splice site downstream of the Sm binding site releasing the lariat-intron and *TER1* as products. Execution of both transesterfication reactions and joining of an artificial second exon to *TER1* 3' end inhibited *TER1* maturation. However, the budding yeast pre-*TLC1*

does not contain an intron, making it unlikely that a similar mechanism is responsible for *TLC1* maturation.

Chapter 2

Materials and Methods

2.1 Materials

2.1.1 Enzymes and chemicals

Modifying enzymes were obtained from New England Biolabs, Roche, Promega or Invitrogen. Standard laboratory reagents were purchased mainly from Sigma and Fluka. Yeast media was obtained from DB Biosciences and Formedium.

2.1.2 Bacterial and yeast culture media

E. coli cultures were grown in Luria-Bertani medium (LB) (10g/l bacto-tryptone, 5g/l bacto-yeast extract, 10g/l NaCl), supplemented with ampicillin (Amp) to 100μg/ml where needed for selective pressure in the propagation of plasmids.

Standard growth and handling techniques were employed for the propagation of *S. cerevisiae*. Yeast was grown at 25°C in either rich media (10g/l yeast extract, 20g/l peptone and 20g/l glucose for YPD or 20g/l galactose and 10g/l sucrose YPGalSuc) or minimal media (6.8g/l yeast nitrogen base with amonnium sulphate, 0.69g/l complete amino acid or dropout powder and 20g/l glucose for SD or 20g/l galactose and 10g/l sucrose SGS). For agar plates 2% (w/v) bacto-agar was added prior to autoclaving at 15lb/sq.inch for 20 min.

2.1.3 Buffers frequently used

1x TE	10 mM Tris-HCl pH 7.6, 1 mM EDTA
10x TBE	0.9 M Tris-borate pH8.3, 20 mM EDTA
1x PBS	140 mM NaCl, 3 mM KCl, 8 mM Na_2HPO_4, 1.4 mM KH_2PO_4, 20 mM $MgCl_2$
20xSSC	3 M NaCl, 0.3 M tri-sodium citrate pH 7.0

2.1.3 Yeast strains

Strain	Genotype	Origin/ Reference
BY4741	*MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0*	Euroscarf
Air1-HTP	*MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; air1-HTP::K.I.URA3*	This study
Air2-HTP	*MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; air2-HTP::K.I.URA3*	This study
Air2-TAP	*MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; air2-TAP::HIS3*	Euroscarf
Nrd1-HTP	*MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; nrd1-HTP::K.I.URA3*	This study
Nab3-HTP	*MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; nab3-HTP::K.I.URA3*	This study
Trf4-HTP	*MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; trf4-HTP::K.I.URA3*	This study
Nop58-HTP	*MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; nop58-HTP::K.I.URA3*	(Granneman et al., 2009)
RRP9-HTP	*MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; rrp9-HTP::K.I.URA3*	(Granneman et al., 2009)
Gal::nrd1	*MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; NAT::GAL10::3HA::nrd1*	This study
Gal::nab3	*MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; NAT::GAL10::3HA::nab3*	This study
Gal::nrd1 Gal::nab3	*MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; NAT::GAL10::3HA::nrd1; KAN::GAL10::3HA::nab3*	This study
Gal::sen34	*MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; NAT::GAL10::3HA::sen34*	This study
BY4741 URA3+ p427Fui1	*MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; URA3 K. lactis [pRS427 Fui1]*	This study
trf4Δ	*MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; trf4::KAN*	Euroscarf
trf5Δ	*MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; trf5::KAN*	Euroscarf
rrp6Δ	*MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; rrp6::KAN*	Euroscarf
Gal::nrd1CIDΔ	*MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; NAT::GAL10::3HA::nrd1CIDΔ*	This study
pap1-5	*MATa; ade2-1; his3-11; 15 trp1-1; ura3-1; leu2-3, 112; pap1::Leu2; [pap1-5]*	(Minvielle-Sebastia et al., 1994)
pap1-5 trf4Δ	*MATa; ade2-1; his3-11; 15 trp1-1; ura3-1; leu2-3, 112; pap1::Leu2; trf4Δ::KAN; [pap1-5]*	L. Milligan personal communication
rna14-1	*MATå; ade2-1; his3-11; leu2-3, 112; trp1-1; ura3-1; rna14-1*	(Minvielle-Sebastia et al., 1991)
rna15-2	*MAT**a;** ade2-1; his3-11; leu2-3,112; trp1-1; ura3-1; rna15-2*	(Minvielle-Sebastia et al., 1991)

sen1-1	*MATå; leu2-3, 112; ura3-52; pep4-3 sen1-1*	(Winey and Culbertson, 1988)
nrd1-102	*MATå; ura3Δ0; his3Δ1; leu2Δ0; met15Δ0; LYS2; nrd1Δ::KAN [nrd1-102] LEU*	(Conrad et al., 2000)
nab3-11	*MATå; ade2; can1-100; his3-11, 15; leu2-3, 112; trp1-1; ura3-1; nab3-11*	(Conrad et al., 2000)
BY4741 +pRS1191 WT	*MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0 [pBS1191] URA3*	This study
BY4741 +pRS1191 U6 59-72Δ	*MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0 [pRS1191 U6 59-72Δ] URA3*	This study
Gal::nrd1 +pRS1191 U6 59-72Δ	*MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; NAT::GAL10::3HA::nrd1; [pRS1191 U6 59-72Δ] URA3*	This study
Gal::nab3 +pRS1191 U6 59-72Δ	*MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; NAT::GAL10::3HA::nab3; [pRS1191 U6 59-72Δ] URA3*	This study
trf4Δ +pRS1191 U6 59-72Δ	*MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; trf4::KAN; [pRS1191 U6 59-72Δ] URA3*	This study

2.1.4 Oligonucleotides

Name	**Sequence 5'-3'**	**Reference**
	Oligonucleotides for strain construction	
WW103Air1 HTP F	ATAAGAGGGGCCGATCATCTTCTAACAAAAGCCAAAGAAAT GGCCGTTATGAGCACCATCACCATCACC	This study
WW102 Air1 HTP R	GTTTGCTGCAATGAGAATGGAAAAAAAATTAAAAAAACTCA CATATAATCTACGACTCACTATAGGG	This study
WW109 Air2 HTP F	ACTACAATTCTTATCAACCATATAGGAGTGGTACCTTGGGGA AAAGAAGAGAGCACCATCACCATCACC	This study
WW108 Air2 HTP R	AAAATATAATGTAACCAAGAACAGCTTGTTAAAGGGCTTCCT ATTTAAAGTACGACTCACTATAGGG	This study
WW138 Nrd1 HTP F	ATTCTTTGATGAATATGCTTAACCAACAGCAGCAGCAACAAC AACAAAGCGAGCACCATCACCATCACC	This study
WW139 Nrd1 HTP R	GGTAGATTAGTTTTATGTACTATGAGCAAATAAAGGGTGGAG TAAAGATCTACGACTCACTATAGGG	This study
WW132 Nab3 HTP F	CTGGCAATAATGTTCAAAGTCTATTAGATAGTTTAGCAAAAC TACAAAAAGAGCACCATCACCATCACC	This study
WW133 Nab3 HTP R	AGAATTCAAGTATAATGTACAAGAAATGGAAAAGATTGAAA AAAGGGAGTTACGACTCACTATAGGG	This study
WW142 Trf4 HTP F	CTGTCTCTAGCGAAGATGATGATGAAG ATGGATATAATCCTTATACCCTTGAGCACCATCACCATCACC	This study
WW143 Trf4 HTP R	GTACACAGTGATGTACAGTTCAGTGCATC ATTTAAACAAAAAGGCACATATACGACTCACTATAGGG	This study
WW156 GAL Nrd1 F4	GAGTTACAGG AAAGGAACCG GAAAGCAACA AACATACTAAACATCCCATAGAATTCGAGCTCGTTTAAAC	This study
WW157 GAL Nrd1 R3	AATGATTCCAAGGTAGCTACAAAATTTTGAAAATCGTCGTCC TGCTGCATGCACTGAGCAGCGTAATCTG	This study
WW158 GAL Nab3 F4	TACGTTTTTC CCGGACTGTC CTTTCATAAT ATAATAACCA TCTGCAAGCCGAATTCGAGCTCGTTTAAAC	This study
WW159 GAL Nab3 R3	TCAGGTGAAGGAATATCTTGAACATCACTGTTATGGTTTTCA TCTGACATGCACTGAGCAGCGTAATCTG	This study
WW231 GAL sen34 F4	ATAGCTATTAACATATATCAAAAAGAACGGCAAAAGGCGAG GAGGTTTTTGAATTCGAGCTCGTTTAAAC	This study
WW163 GAL sen34 R3	CATTTCCTTAGAAGTTTGATGTGATCTATGTCAAATACTAGC GGTGGCATGCACTGAGCAGCGTAATC	This study

	Oligonucleotides for strain testing	
KanB	CTG CAG CGA GGA GCC GTA AT	Wach 1994 Yeast
MX6 F	CCT CGA CAT CAT CTG CCC AGA T	M. Kos Tollervey lab
MX6 R	TGC AGC GAG GAG CCG TAA	M. Kos
Air1 C	GCAGTCAACAAGGAACAAAAATAGT	Euroscarf
Air1 D	TGACAACTGGTTACTTCCCTAAGAC	Euroscarf
Air2 C	AAATACAAAAGTAAACGTCTCGTCG	Euroscarf
Air2 D	AAATTGAGTTGCATAGATGTTAGCC	Euroscarf
Nrd1 A	ATATCAAAAGTCAGTACTCGGCAAC	Euroscarf
Nrd1 B	TATCAATGTGATCAAGTGCGTAAGT	Euroscarf
Nrd1 C	ATTCTGAATAACTCCAGAAAACACG	Euroscarf
Nrd1 D	TTTACCTTCATGGTACTGGATAAGC	Euroscarf
Nab3 A	AGTGTAACCCTGAATTGTTGAAGAG	Euroscarf
Nab3 B	AAGTTCGACCTCTTTATCTTTGGTT	Euroscarf
Nab3 C	CTCCTCCTCAAACAAACTATTACCA	Euroscarf
Nab3 D	CCAAATAGCATTTGATAAGGAAGAA	Euroscarf
Trf4 A	ACGCTCTTGAATTTAGAATAGCTGA	Euroscarf
Trf4 B	TAAGGATTTCCTTTTGCTTTTCTTT	Euroscarf
Trf4 C	AAGAGATTTCAAGGATGAAAGAGGT	Euroscarf
Trf4 D	ATTTCCTAATGATACCACTGCTGAG	Euroscarf
Trf5 A	GAAGGCACCTTACTAGACCATTGTA	Euroscarf
Trf5 D	AGGTTCTTTCAAATTATGTGTCTCG	Euroscarf
Rrp6 A	TGACAGAACCATTTCATGTTCAATA	Euroscarf
Rrp6 D	ATGTGAAGAAAAGAATTCCTGACAC	Euroscarf

	Linker and Oligonucleotides for CRAC	
MiRCat linker	rAppTGGAATTCTCGGGTGCCAAGG/ddC	(Granneman et al., 2009)
MirCat RT oligo	CCTTGGCACCCGAGAATT	(Granneman et al., 2009)
MiRCat to Solexa PCR oligo	CAAGCAGAAGACGGCATACGACCTTGGCACCCGAGAATTCC	S. Granneman personal communication
RL5 linker	invddT-GTTCArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrC	(Granneman et al., 2009)
L5a	invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrArC	S. Granneman personal communication
L5c	invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrGrA	S. Granneman personal communication
L5d	invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrArCrArArGrC	S. Granneman personal communication
L5e	invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrCrArCrArGrC	S. Granneman personal communication
5’ Solexa PCR	AATGATACTGCGACCACCGACAGGTTCAGAGTTCTACAGTCCGA	(Granneman et al., 2009)
M13 F	GTAAAACGACGGCCAGT	
M13 R	AACAGCTATGACCATG	

	Oligonucleotides for riboprobe substrates	
WW255 T7 HPF F as	GGATCCTAATACGACTCACTATAGGGAGAGGACTCCGGTTCATCTTCTGCC	This study
WW256 HPF R as	AGAGGCAGTGGAAGCCGAT	This study
WW278 T7 HPF F mRNA	GGATCCTAATACGACTCACTATAGGGAGAGGAAGAGGCAGTGGAAGCCGAT	This study
WW279 HPF R mRNA	CTCCGGTTCATCTTCTGCC	This study

WW304 T7 CAF17 F as	GGATCCTAATACGACTCACTATAGGGAGAGGAATTACTCCAC ATTTCCACAATCCT	This study
WW305 CAF17 R as	GCTGTCAGCTCTTGTCCTACATAA	This study
WW320 CAF17 F mRNA	GGATCCTAATACGACTCACTATAGGGAGAGGAGCTGTCAGCT CTTGTCCTACATAA	This study
WW321 CAF17 R mRNA	ATTACTCCACATTTCCACAATCCT	This study

	Hybridisation oligo probes	
004 20S	CGGTTTTAATTGTCCTA	
041 5S	CTA CTC GGT CAG GCT C	
040 5.8S	GGAAUACCAAGGGGCGCdAdA	David 96
U6	ATCTCTGTATTGTTTCAAATTGACCAA	(Tong and Boone, 2006)
U4	AGGTATTCCAAAAATTCCC	(Tong and Boone, 2006)
304 tRNA Pro	ACCCAGGGCCTCTCG	(Kufel et al., 2002)
305 tRNA Trp	AACCTGCAACCCTTCGA	(Kufel et al., 2002)
306 tRNA Leu exon	GCA TCT TAC GAT ACC TG	(Kufel et al., 2002)
307 tRNA Leu intron	CAC AGT TAA CTG CGG TC	(Kufel et al., 2002)
308 tRNA Phe intron	AAC TTG ACC GAA GTA TTT C	(Kufel et al., 2002)
320 tRNA Lys mature	ATCCTTGCTTAAGCAAATGCGC	(Kufel et al., 2002)
327 tRNA Lys intron	CTCTACCAACTGAGCTAAC	(Kufel et al., 2002)
W327 tRNA Phe exon	GGATCGAACACAGGACCTCCA	This study
329 tRNA Ser intron	AGCCGAACTTTTTATTCCATTCG	(Kufel et al., 2002)
330 tRNA Ser mature	AGCCCAAGAGATTTCGAGTCTCTCG	(Kufel et al., 2002)
499 TSA1	GGAGTATTCGGAGTCAGTGGAGGCGAAAAGAACT	(Houseley and Tollervey, 2006)
W343 TLC1	GTGTGGTGATGGTAGGCTTCCCATGG	(Chapon et al., 1997)

	Oligos for site directed mutagenesis	
WW340 U6 d59-72 F	GGTCAATTTGAAACAATACAGAGATGATCATAAGGATGAAC CG	This study
WW341 U6 d59-72 R	CGGTTCATCCTTATGATCATCTCTGTATTGTTTCAAATTGACC	This study

	5' RACE oligos	
WW338 HPF1 as RACE	CCTCTGAGGCCCCTGAAGCCACAA	This study
WW339 CAF17 as RACE	CGAGC CAACCCTCAATCCATTTACC AATAA	This study

2.1.5 Plasmids

Plasmid	**Description**	**Reference/Remarks**
pFA6a-KanMX6	Used as a template for PCR amplification of the Kan deletion cassette	(Longtine et al., 1998)
pFA6a-NatMX6	Used as a template for PCR amplification	(Hentges et al., 2005)

	of the Nat deletion cassette	
pFA6a-NatMX6-*PGAL1*-3HA	Used as a template for PCR amplification of the N-terminal NAT::*GAL*::3HA tag	(Hentges et al., 2005)
pFA6a-KanMX6-*PGAL1*-3HA	Used as a template for PCR amplification of the N-terminal KAN::*GAL*::3HA tag	(Longtine et al., 1998)
pRS1539 HTP-URA	Used as a template for PCR amplification of the C-terminal 6His-Tev-ProtA tag	(Granneman et al., 2009)
pBS1191	U6 gene 500nt up and downstram; *CEN*; URA3	(Luukkonen and Seraphin, 1998)
pBS1191 U6 59-72Δ	Deletion of nt 59-72 in U6 gene in pBS1191	This study
pPM1	T7-`RPR1-T3	Phil Mitchell personal communication
pRS427 Fui1	Uridine permease Fui1 under its own promoter	Alex Tuck personal communication

2.1.6 Radiolabelled compounds

All radionuclides were purchased from Perkin Elmer: [γ-32P] ATP (6000 Ci/mmol), [α-32P] UTP (3000 Ci/mmol).

2.1.7 Antibodies

Antibody	**Description (conc. used)**	**Supplier/ Reference**
PAP Peroxidase-anti-peroxidase	Rabbit (1:5000)	Sigma UK
Anti-HA-HRP (F-7)	(1:1000)	Santa Curz
Anti-TAP	Rabbit (1:5000)	Open biosystems
Secondary-HRP abs	Various species (1:30 000)	Sigma

2.2 Methods

2.2.1 Bacterial and yeast techniques

2.2.1.1 Inoue competent cells

A culture of *E.coli* DH5α was grown in LB at 37°C to OD600 0.55 and transferred to an ice water bath immediately. Cells were harvested at 2500 g for 10 min at 4°. Pellets were resuspended in 80 ml ice-cold transformation buffer (55 mM $MnCl_2$, 15 mM $CaCl_2$, 250 mM KCl, 10 mM PIPES pH 6.7) and centrifuged as before. Cells were resuspended in 20 ml ice cold transformation buffer with 1.5 ml DMSO and incubated on ice for 10 min. Aliquots of 100 µl were frozen in liquid nitrogen and stored at -80°C until needed.

2.2.1.2 Plasmid transformation into *E.coli*

Chemically competent E.coli were thawed on ice. Approximately 10 ng of plasmid DNA was added to 50 µl cells and they were incubated on ice for 15 min. Heat shock was performed for 60 sec at 42°C in a water bath, afterwards cells were immediately transferred back to ice and incubated for 5 min. 1 ml LB was added to the cells and transferred to a 37°C incubator for 60 mins. Cells were pelleted at 2500 g in a micro centrifuge, resuspended in 100 µl LB and spread onto LB agar containing amplicillin. Transformations were grown over night at 37°C.

2.2.1.3 Yeast transformation

Yeast strains were transformed using a lithium acetate method (Gietz) to generate conditional alleles, for gene disruption, for epitope tagging or to introduce a plasmid. An overnight yeast culture was diluted to OD_{600} of 0.3 (5 ml per transformation) and grown to OD_{600} of 1.0 ($2x10^7$ cells/ml). After harvesting, cells were washed with 5 ml sterile water and centrifuged again. Then 250 µl PEG 4000 solution (40%), 36 µl LiOAc (1M), 50 µl salmon sperm carrier DNA (50 mg/ml in TE, Sigma) and 1-5 µg transforming DNA was added to the pellet. The cells were resuspended by vortexing

and heats hocked for 40 min at 42°C. Then the cells were centrifuged, resuspended in 200 µl sterile water and plated onto YPD, YPGalSuc or SD plates and incubated at 25°C. For selection with an antibiotic resistance marker cells were replica-plated onto selective media after 24 h of growth.

2.2.2 Recombinant DNA techniques

2.2.2.1 Automated DNA sequencing

Sequencing reactions were performed using the Big Dye sequencing ready reaction (ABI). A sequencing reaction contained 2 µl Big Dye mix, 2 µl 5x Big Dye buffer (20 mM Tris pH 8.0, 5 mM $MgCl_2$), 200 ng template DNA, 1 µl primer (1.6 pmol/µl) and water to 10 µl. The sequencing reaction was run as follows:

Step 1 96°C	1 min	
Step 2 96°C	10 sec	
Step 3 50°C	5 sec	
Step 4 60°C	2 min	Return to step 2 for a further 24 cycles

Sequencing reactions were analysed by THE GENE POOL sequencing service of the University of Edinburgh.

2.2.2.2 Polymerase chain reaction (PCR)

PCR was used for amplification of tagging or deletion cassettes from plasmids, regions of the yeast genome for mobilisation and/or mutagenesis of cassettes and for amplification of DNA fragments, generation of *in vitro* transcription templates and identification of recombinant clones (colony PCR). All applications followed the general protocol described below, or modifications upon it.

Amplification of DNA for cloning purposes utilised the high fidelity enzyme Phusion. A typical 50 µl reaction contained 10 µl 5x High Fidelity reaction buffer, 1 µl primer (sense and antisense 10 pmol/µl each), 1 µl dNTP mix (10 mM of each

dNTP), 0.02 U Phusion polymerase and varying amounts of template DNA. For amplifications from genomic DNA templates the reaction was supplemented with DMSO to 3% final concentration. The final volume was obtained by addition of sterile water to 50 µl. A typical reaction profile used is:

Step 1 98°C	30 sec	
Step 2 98°C	10 sec	
Step 3 50°C	30 sec	
Step 4 72°C	30 sec/kb	Return to step 2 for a further 34 cycles
Step 5 72°C	7 min	

Colony PCR was performed using Taq polymerase (NEB) under similar conditions. A single yeast colony was heated to 95°C for 5 min in 30 µl 0.2 % SDS and briefly centrifuged. For each reaction 0.5 µl of the supernatant was used.

2.2.2.3 Plasmid preparations

Over-night cultures of single *E.coli* colonies were grown in LB media supplemented with 0.1 mg/ml Ampicillin at 37°C. 3 ml of a saturated culture were harvested and plasmids were purified using a plasmid mini prep kit (Quiagen) according to the manufacturers instructions.

2.2.2.4 Site directed mutagenesis

In order to introduce point mutations or deletions into plasmid DNA, site directed mutagenesis PCRs were performed. Oligonucleotides bearing the mutation were constructed according to the specification given in the manual of the site directed mutagenesis kit (Stratagene). PCR reactions were performed containing 50 ng plasmid DNA, 125 ng forward and reverse oligos, PCR reaction buffer and Pfu turbo polymerase (2.5 U/µl, Promega); a negative control without oligos was used. PCR reactions with 18-25 cycles were carried out as described in 2.2.2.2, annealing times were however extended to 1 min at 55°C and elongations were carried out at 68°C

for 1 min per kb DNA. The PCR products were digested with 1 µl DpnI (xU /µl, NEB) for 2 hours at 37°C to remove the template plasmid. Approximately $1/10^{th}$ of the digestion reactions was transformed into *E.coli* DH5alpha according to 2.2.1.2. Transformants were screen for correct mutations by sequencing.

2.2.3 DNA techniques

2.2.3.1. Preparation of total yeast DNA and phenol chloroform extraction

5 ml of an overnight yeast culture were harvested by centrifugation and resuspended in 200 µl breaking buffer (2% v/v TritonX 100, 1% v/v SDS, 100 mM NaCl, 10 mM Tris pH 8.0, 1 mM EDTA pH 8.0). Then 200 µl phenol:chloroform:isoamylalcohol (25:24:1) and 200 µl glassbeads were added to the cell suspension and vortexed for 5 min. The mixture was centrifuged briefly, 200 µl TE were added. For separation of the phases the mixture was cenrifuged for 5 min at 20000 g. The upper, aqueous phase was extracted with an equal amount of chloroform:isoamylalcohol (24:1) and then added to 1 ml EtOH (96%), 40 µl NaOAc (3M, pH 5.2) for precipitation. The DNA pellet was washed with 70% EtOH, allowed to air-dry and resuspended in 50 µl TE.

2.2.4 RNA techniques

2.2.4.1 Yeast total RNA prep with GTC

RNA was extracted as previously described (Tollervey 1987). Approximately $2x10^7$ of exponentially growing cells (OD_{600} of 0.2-0.8) were harvested by centrifugation, washed with sterile water and collected as before. Cell pellets were frozen in liquid nitrogen and stored at -80°C until needed. All steps were performed on ice or at 4°C except if stated otherwise. Cell pellets were resuspended in 40 µl GTC phenol mix (4M guanidine thiocyanate, 0.05M Tris pH 8.0, 0.01M EDTA pH 8.0, 2% sarkosyl, 1% ß-mercaptoethanol, 50% phenol), 50 µl of zirconia beads (Thistle) were added

and vortexed for 5 min. Then an additional 600 µl GTC phenol mix was added, the mixture was vortexed briefley and incubated for 10 min at 65°C to denature the RNA followed by a 10 min incubation on ice. Afterwards 160 µl NaOAC Mix (100 mM NaOAc, 1 mM EDTA, 10 mM Tris-HCl pH 8.0) and 300µl chloroform were added. For phase separation the mixture was centrifuged for 20 min at 20000 g in a chilled microcentrifuge. The aqueous phase was extracted with an equal volume of phenol:chloroform:isoamylalcohol (25:24:1), chloroform:isoamylacohol (24:1) and then precipitated with 2.5 volumes EtOH. The pellet was washed with 70% EtOH, briefly air-dried and resuspended in H_2O. This procedure was scaled up or down to the experimental requirements.

2.2.4.2 Hot phenol RNA extraction from yeast

25 ml of exponentially growing cells at OD_{600} 0.5 were harvested by centrifugation at 850 g and washed once with H_2O. Cell pellets were resuspend in 400µl AE (50mM NaOAc pH 5.3, 10mM EDTA; adjusted to pH 5.3 with acetic acid) and 10% SDS added to 0.9 % final conc. An equal volume of Tris-buffered phenol (440µl) (pH 7.5; Sigma) was added and vortexed to mix. The mixture was incubated for 45 min at 65°C in a shaking heating block at 1000 rpm, cooled to RT on ice and then spun for 5 min in a table top microfuge at 20000 g and RT. 400 µl aqueous phase were extracted with an equal volume phenol: chloroform (5:1; pH 4.5; Ambion), vortexed and left at room temperature for 5 min before spinning for 5 min in a microfuge as before. The aqueous phase was extracted with an equal volume of chloroform/isoamylalcohol as before and RNA precipitated with 2.5 volumes of 100% EtOH. The RNA was collected by centrifugation and washed as in 2.2.4.1

2.2.4.3 poly(A)$^+$ prep

Polyadenylated and oligoadenylated RNA were isolated from total yeast RNA with a polyA tract mRNA isolation Kit IV (Promega) according to the manufacturer's protocol. To recover oligoadenylated RNAs, the stringency of the washes was lowered from 0.1xSSC to 0.2xSSC.

2.2.4.4 Pulse-chase labelling with 4-thiouracil and nascent RNA isolation

Cells expressing the uridine permease Fui1 on a plasmid (pRS427 Fui1) and any URA3 resistance cassette were grown SD –Ura -Leu media to OD_{600} 0.5. 25ml cells were harvested per time point by addition of the cell suspension to 25 ml EtOH on dry ice (Kos and Tollervey, 2010). Cells were collected by centrifugation and washed, if needed repeatedly with H_2O to remove any possible traces of $(NH_4)_2SO_4$. The first sample was taken prior to 4-thiouracil addition (1M 4-thiouracil in DMSO; 20µl per 100 ml culture). The 4-thiouracil pulse was continued for 10 min, then cells were collected by filtering through a (0.8µm AAWP Nitrocellulose filter, Millipore), washed once with pre-warmed SD –Leu and resuspended by immersing the filter in fresh pre-warmed SD –Leu media (chase). Cells were harvested as described above over a time course and RNA was extracted by the hot phenol extraction method in 2.2.4.2.

Equal amounts of total RNA (usually 100-200µg) were used for subsequent modification and biotinylation.

Total RNA was diluted into 250 µl 1x TE2 (50 mM Tris-HCl pH 7.8, 1 mM EDTA) and incubated with 20µl TECP agarose slurry (Pierce) for 2 hr at 16 °C with gentle agitation. The beads were collected by brief centrifugation and the supernatant was used for biotinylation. 25 µl HPDP-biotin (4 mM in dimethylformamid) were added to the RNA and incubated for 3 h at RT in the dark. In order to remove excess HPDP-biotin, the RNA was extracted once with chloroform/isoamylalcohol and precipitated with 2.5 volumes of EtOH and $1/10^{th}$ volume of NaOAc as before. RNA was resuspended in RBS_{100} (10 mM Tris pH 7.5, 100 mM NaCl, 2.5 mM $MgCl_2$, 0,4% TritonX-100) and bound to 50 µl streptavidin magnetic beads slurry (Roche) for 30 min at 4˚C. Beads were pre-blocked in RBS100 with 200µg/ml glycogen (Roche) for 30 min at RT on a rotating wheel. The beads were washed 3x with 500µl $RBS_{100,}$ 2x with 500µl TEN_{1000} (10mM Tris-HCl pH 7.5, 1mM EDTA pH8, 1M NaCl), once with H_2O and nascent RNAs were eluted 2x in 100 µl by reduction with 0.1 M ß-mercaptethanol for 5 min at RT. RNAs were precipitated with EtOH,

NaOAc and glycogen as described before over-night at -20˚C and analysed by gel electrophoresis and northern blotting.

2.2.4.5 RNA gel electrophoresis and northern blotting

Polyacrylamide and agarose gel electrophoresis were performed as has been previously described (Maniatis et al., 1982, revised 1989).
Sequencing gels contained 12% acrylamide-bis-acrylamide (18:1), 8.3 M urea and 1x TBE. They were 40 cm long and 0.1 mm thick. Prior to electrophoresis RNA samples were denatured for 3 min at 95°C in FA loading dye (47% formamide, 10 mM EDTA, 0.025% bromphenol blue and 0.025% xylene cyanol) and snap-chilled on ice. Following electrophoresis gels were dried on Whatman 3MM paper before exposure to a phosphoimager screen.
Low molecular weight RNAs were separated on 8 or 10% polyacrylamide TBE 8M urea gels (15 cm long and 1.5 mm thick). Following electrophoresis gels were stained with SYBRsafe (1:50000 in 0.5x TBE, Invitrogen), scanned on the Fujifilm FLA-5100 phosphoimager (532 nm) and transferred to nylon membrane (Hybond N+, GE) by electroblotting (15V) in 0.5x TBE over night at 4°C. RNA was crosslinked to the membrane by UV light (120 mJ/cm^2) in a Stratalinker.
High molecular weight RNAs were separated on 1% agarose BPTE gels, with 1x BPTE running buffer for 16 hours at 45 V. Prior to loading total RNA was was denatured in glyoxal mix (60% DMSO, 20% glyoxal, 1xBPTE, 0.05% glycerol containing 20 μg/ml ethidium bromide, 200 μg/ml for the marker; RNA:glyoxal mix 1:5) for 1 hour at 55°C. Following electrophoresis the gel was scaned on the phosphoimager, fixed for 20 min in 75 mM NaOH, washed in neutralising solution (1.5 M NaCl, 0.5 M Tris pH 7.5) for 20 min and transferred to Hybond N+ in 6xSSC by capillary force over-night. RNA was crosslinked to the membrane as before.

2.2.4.6 5’ end labelling of oligoprobes

Oligonucleotides were 5’ labelled for northern hybridisation (5 pmol oligo) or primer extension (2.5 pmol oligo) in a 10 μl reaction containing 1x PNK raction buffer, 10

mM DTT, 2 µl (20 µCi) [γ^{32}P]ATP and 10 U polynucleotide kinase (T4 PNK, NEB) at 37°C for 1 hour. Afterwards PNK was heat inactivated for 5 min at 65°C. Unincorporated nucleotides were removed by spinning of the probe through mini quick oligo spin columns (Roche).

2.2.4.7 *in vitro* transcription of riboprobes

RNA transcripts were generated from 100 ng linearized plasmid or 50 ng transcription template PCR products. A 25 µl reaction also contained 0.6 µl 100 µM UTP, 1 µl 10 mM CTP/GTP/ATP mix, 2 µl 100 mM DTT, 0.2 µl 10 mg/ml BSA, 1x transcription buffer, 5 µl (50 µCi) [α-^{32}P]UTP and 1 µl T3 or T7 RNA polymerase (HC 80 U/µl, Promega) and was incubated for 1 hour at 37°C. Unincorporated nucleotides were removed by spinning the probe through mini quick spin columns (Roche).

2.2.4.8 Hybridisation of northern blots

Oligoprobes were hybridised with the membranes over night at 37°C in oligo-hyb buffer (7% SDS, 170 mM Na_2HPO_4, 80 mM NaH_2PO_4 0.5 mM EDTA). Prior to addition of the labelled oligo, membranes were pre-hybridised in oligo-hyb for 1 h at 37°C. After hybridisation membranes were rinsed once with 6xSCC, washed once with 6xSCC at RT for 10 min and washed once under higher stringency with pre-warmed 2x SSC, 0.2% SDS at 37°C for 10 min. Membranes were dried with tissue paper, wrapped in saran wrap and exposed to a phosphoimager screen. Images were scanned by a Fujifilm FLA-5100 phosphoimager and images quantified with AIDA software where appropriate.

Riboprobes were hybridised on membranes over night at 65°C in Ultra-Hyb sensitive hybridisation buffer (Ambion). Pre-hybridisation of the membrane in Ultra-Hyb was carried out while the oven was heating to 65°C and continued for further 30 min at 65°C prior to addition of the purified probe. After hybridisation membranes were rinsed twice with 6xSSC and washed once with 6xSSC at RT for 15 min. Two high

stringency washes were carried out with pre-warmed 0.2xSSC, 0.2% SDS at 65 °C for 15 min. Membranes were dried and exposed as described before.
Probes were stripped off the membrane by washing the membranes twice in boiling 0.1xSSC, 0.1% SDS and used for further hybridisations.

2.2.4.9 Primer extension analysis

0.1 pmol end labelled oligo and 4 µg of total yeast RNA per reaction were denatured in the presence of 2 M betain at 98°C for 3 min and snap-chilled on ice. A reaction mixture containing 1x first strand buffer, 1 µl (40 U/µl rRNasin, Promega), 5 mM DTT and 2 mM dNTP mix was prepared and pre-warmed to 50°C. The denatured RNA-oligo mix was added to the pre-warmed reaction and incubated for 3 min at 50°C. 1µl Super Script III reverse transcriptase (5 U/µl, Invitrogen) was added and reactions carried out for 30 min at 50°C. Afterwards the reverse transcriptase was heat inactivated by incubation at 65°C for 20 min. FA loading buffer was added to the reaction 1:1 and 1/20 of the reaction was analysed on a polyacrylamide sequencing gel. Primer extension products were compared to a radioactively labelled size marker.

2.2.5 Protein and immunological techniques

2.2.5.1 Yeast protein isolation

Approximately 1 OD of cells in logarithmic growth ($2x10^7$ cells/ml) were pelleted, resuspended in 15 µl 2 M NaOH with 80 mM DTT and incubated on ice for 10 min. 15 µl of 50 % TCA were added, vortexed to mix and incubated for a further 10 min on ice. The mixture was centrifuged for 2 min at 20000g at RT, the supernatant removed and the pellet was resuspended in 1 ml cold acetone and precipated again. The pellet was air-dried briefly and resuspended in 20 µl sample buffer (100 mM Tris pH 8.8, 2% SDS, 0.1% bromophenol blue, 10% glycerol, 100 mM DTT) or 1x NuPAGE sample buffer (Invitrogen).

2.2.5.2 SDS polyacrylamide gel electrophoresis (PAGE)

Proteins were separated on Tris-glycine polyacrylamide SDS gels as described by (Maniatis et al., 1982, revised 1989) or on NuPAGE gradient gels (Invitrogen). Proteins in sample buffer (see above) were denatured before loading by heating to 98°C for 5 min and collected by brief centrifugation. Gels were run according to the manufacturer's protocol.

2.2.5.3 Western blotting

Following PAGE, proteins were transferred to nitrocellulose membrane (Hybond C, GE) in a wet blot apparatus (Bio Rad). The gel was assembled on the nitrocellulose membrane between two sheets of Whatmann 3MM paper on each side, soaked in wet transfer buffer (Invitrogen, supplemented with 15% methanol). Wet transfer was performed for 1.5 hours at 100 V. After transfer the membrane was blocked with 5% low-fat dried milk (w/v) in TBS (10 mM Tris-HCl pH 8.0, 250 mM NaCl) 0.1% Tween 20.

For protein detection the blot was decorated with different antibodies (see table in 2.1.7 for details). All antibodies were diluted as stated in 5% milk TBS 0.1 % Tween 20. Primary antibodies were incubated with the membrane over night at 4°C. Secondary horseradish peroxidase coupled (HRP) antibodies were incubated for 1 hour at RT. If primary antibodies were directly coupled to HRP, they were incubated for 2 hours at RT instead. All incubation steps with antibodies were followed by three washes with TBS 0.1% Tween for 10 min at RT. Proteins were detected using the HRP substrate, enhanced chemiluminescence (ECL) Kit (Pierce), following the manufacturer's instructions.

2.2.6 RNA protein crosslinking techniques (CRAC)

2.2.6.1 UV crosslinking, extract preparation, IgG binding and TEV cleavage

Yeast strains carrying genomically encoded C-termial HTP-tagged proteins were grown to OD600 0.5 (1 l per experiment), harvested by centrifugation, washed in cold PBS and the pellets resuspended in one volume PBS. The yeast suspension was spread on a petri dish and chilled on ice. Cells were crosslinked *in vivo* with four blasts of 400 mJ/cm2 UV light (254 nm) in a Stratalinker. Afterwards cells were collected by centrifugation, pellets were frozen in liquid nitrogen and stored at -80°C.

All following steps were carried out on ice unless otherwise stated. For extract preparation the pellets were quickly thawed in one's hand and resuspended in 1 volume TMN500 (50 mM Tris-HCl pH 7.8, 500 mM NaCl, 1.5 mM $MgCl_2$, 0.1% NP40) containing 5 mM ß-mercaptoethanol (ß-ME) and complete EDTA-free protease inhibitors (PIs, Roche, 1 tablet per 50 ml). Cells were lysed by vortexing with 2.5 volumes Zirconia beads (Thistle) for 1 min, chilled on ice 1 min and repeated 4 times. Further 3 volumes of TMN500, 5 mM ß-ME and PIs were added and the lysate was cleared from Zirconia beads and cell walls by centrifugation at 4600g in a chilled Falcon tube centrifuge. The supernatant was transferred to Eppendorf tubes and the lysate cleared from chromatin and unsoluble membranes by centrifugation at 20,000g and 4°C. The cleared lysate was added to 1/10 V packed IgG sepharose beads equillibrated in TMN500 and incubated on a shaking platform at 4°C for 2 hours. Following IgG binding the beads were washed 2x with TMN500 and 3x with TMN150 (50 mM Tris-HCl pH 7.8, 150 mM NaCl, 1.5 mM MgCl2, 0.1% NP40). After the last wash, the beads were resuspended in 600 µl TMN150 with 5 mM ß-ME and 2µl GST-TEV protease (generously supplied by Dr. Sander Granneman and Dr. Simon Lebaron) and proteins were eluted by TEV cleavage in a shaking incubator at 18°C.

2.2.6.2 Partial RNase digestion and Ni affinity purification

In order to generate a 'footprint' of the protein on the RNA, the RNA in the TEV eluate was partially digested with a mixture of RNase A and T1 (RNace it 10 U/µl, Stratagene; here used 1 U per reaction) for 5 min at 37°C. Immediately after the RNase digestion, the mixture was added to 0.4 g guanidinium hydrochloride (final concentration 6 M) to stop the digestion. NaCl and imidazole were added to the reaction to a final concentration of 300 mM and 10 mM, respectively. The digested TEV eluate was then added to 50 µl Ni-agarose beads (Quiagen) equillibrated in wash buffer 1 (WB 1, 50 mM Tris-HCl pH 7.8, 300 mM NaCl, 10 mM imidazole, 6M guanidine-HCl, 0.1% NP40, 5 mM ß-ME, made fresh) and incubated over night at 4°C on a shaking platform. After nickel binding the beads were washed 2x with WB 1 and 3x with PNK buffer (250 mM Tris-HCl pH 7.8, 50 mM MgCl2, 5 mM ß-ME) and proceeded with the enzymatic reaction of the next step or protein RNA complexes were eluted and analysed by western blotting as follows. Following Ni-binding, the beads were washed 3x with wash buffer 2 (WB 2, 50 mM Tris-HCl pH 7.8, 50 mM NaCl, 10 mM imidazole, 0.1% NP40, 5 mM ß-ME) and RNPs were eluted twice with 200 µl elution buffer (10 mM Tris-HCl pH 7.8, 50 mM NaCl, 150 mM imidazole, 0.1% NP40, 5 mM ß-ME) for 5 min at RT. For further analysis proteins were precipitated with 1/10 volume 100% TCA and 20 µg BSA, washed with acetone, resuspended in NuPAGE loading buffer and SDS PAGE was run and western blots preformed as described in 2.2.5.2 and 2.2.5.3.

2.2.6.3 on bead RNA dephosphorylation, radio labelling and linker ligations

If crosslinked RNAs were further analysed, several enzymatic reactions were performed while the crosslinked RNA-protein complexes were still attached to the Ni-agarose. In order to remove the 3' phosphate that the RNase digestion left on the RNA, the RNAs were treated with alkaline phosphatase. Therefore the Ni-agarose was transferred to micro spin columns (Biorad) and resuspended in a 80 µl reaction mixture containing, 8µl TSAP (thermostable alkaline phosphatase, Promega 1 U/µl),

2 µl RNasin (human non recombinant 40 U/µl) and 16 µl 5x PNK buffer (250 mM Tris-HCl pH 7.8, 50 mM $MgCl_2$, 5 mM ß-ME). Dephosphorylation was carried out for 30 min at 37°C. To inactivate TSAP, the beads were washed 3x with WB 1 and 3x with PNK to equilibrate the beads for the next reaction.

Following dephosporylation, a linker was ligated to the 3' end of the RNA. The beads were resuspended in a 80 µl reaction containing 8 µl SOLEXA 3' linker (10 µM), 2 µl RNasin, 2 µl T4 RNA ligase (NEB, 10 U/µl) and 16 µl 5x PNK. The DNA linker (see table oligos) used here has a blocked 3' end and an activated adenosine at the 5' end, such that the ligation can be performed in the absence of ATP. Linker ligation was carried out for 6 hours at 25°C. RNA ligase was inactivated by washes as described before.

To visualise RNPs the RNA was radioactively labelled at the 5' end. The reaction was carried out in 80 µl total volume and contained 4 µl (40 µCi) [γ^{32}P]ATP, 4µl T4 PNK (from phage infected *E.coli,* Sigma, 5 U/µl) and 16µl 5x PNK buffer for 40 min at 37°C. To completely phosphorylate all 5' ends, 1µl 100 mM ATP (lithium-salt, Roche) was added and the reaction proceeded for further 20 min. The enzyme was inactivated and the beads re-equilibrated for the next reaction by washed as described before.

Finally, an RNA linker was ligated to the 5' end of the RNA. A 80 µl reaction contained 2 µl RL5 linker or barcoded L5a, L5c, L5d or L5e (100 µM, IDT), 8 µl ATP (10 mM), 4 µl RNA ligase (NEB, 10 U/µl), 2 µl RNasin and 16 µl 5x PNK buffer. Ligations were carried out over night at 16°C.

2.2.6.4 SDS PAGE, blotting and RNA elution

Following linker ligation the radiolabelled RNA-protein complexes were eluted from the nickel beads as decribed in 2.2.6.2, blotted onto nitrocellulose and detected by autoradiography. According to the autoradiogram, regions containing the RNP were cut from the membrane and the RNA was released from the membrane by Proteinase K digestion. Therefore the membrane was incubated for 2 hour at 55°C in 400 µl WB 2 containing additionally 1% SDS, 5 mM EDTA and 100 µg Proteinase K

(Roche). RNAs were extracted once with phenol chloroform and precipitated with 2.5 volumes EtOH and 20 µg of glycogen (Roche).

2.2.6.5 cDNA sythesis, gelpurification, cloning and sequencing

For cDNA synthesis the precipitated RNA was resuspended in a 10 µl reaction mix containing 1 µl SOLEXA RT oligo (10 µM) and 2 µl dNTPs (5 mM). This mixture was denatured for 3 min at 80°C and snap-chilled on ice. For first strand synthesis 4µl 5x reaction buffer, 1 µl DTT (100 mM) and 1 µl rRNasin (40 U/µl, Promega) were added and the reaction was annealed at 50°C for 3 min. The reaction was started by addition of 1 µl Super Script III (200 U/ µl, Invitrogen) and proceeded for 1 hour at 50°C. The reverse transcriptase was heat inactivated for 15 min at 65°C and th e template RNA was digested for 30 min at 37°C with 2 µl RNase H (5 U/µl, NEB).

To generate libraries for sequencing, PCR reactions were carried out containing the following mix, 5µl 10 LA taq buffer, 0.5 µl LA TaKaRa Taq (5 U/µl, Lonza), 1 µl Solexa PCR oligos forward and reverse (10 µM), 2.5 µl dNTPs (5 mM) and 1 µl of the reverse transcription. The programme was as follows.

95°C 2 min
98°C 20 sec
52°C 20 sec
68°C 20 sec
25 cycles
72°C 5 min

For all the experiments performed 3 of the above PCR reactions were performed per sample, pooled, precipitated with EtOH and ran on a 3% agarose TBE gel (Metaphore agarose, Lonza). DNA in the size range of 80-120 bp was cut out from the gel and purified using Mini Elute Gel Purification Kit (Quiagen) according to the manufacturer's instructions. DNA was eluted in 20 µl H2O. 2 µl of the library were cloned into pCR4 TOPO vector for sequencing (Invitrogen) following the manufacturer's protocol and transformed in to library competent *E.coli* TOP10 (Invitrogen) like described in 2.2.1.2. Over-night cultures of single bacterial clones

were sequenced by THE GENE POOL, University of Edinburgh sequencing service. Coloniy PCRs were performed with T7 and T3 oligos, sequencing was performed with T7 oligo. The remaining library was Solexa sequenced by THE GENE POOL.

2.2.6.6 CRAC bioinformatics

Bioinformatic analysis of the Sanger or Solexa sequencing data was performed in collaboration with Dr. Grzegorz Kudla and was previously described in (Bohnsack et al., 2009; Granneman et al., 2009).

2.3 Frequently used online databases and tools

SGD
Genomic tRNA database
Yeast snoRNA database
ClustalW
Pubmed
Affymatrix IGB
Yeast Deletion Webpages

Chapter 3

UV crosslinking of Air1, Air2, Nrd1, Nab3 and Trf4- Identification of known targets for the nuclear RNA surveillance machinery

3.1 Introduction

Large amounts of RNA get transcribed in the cell by one of three RNA polymerases. Before these transcripts can fulfil their function, usually as part of an RNP complex, they undergo numerous processing and modification steps. Thus proteins associate with and dissociate from the RNA. At every stage of this maturation pathway errors can occur and the cell has established an efficient surveillance system that monitors RNA processing and degrades aberrant RNAs (reviewed by Houseley et al., 2006; Houseley and Tollervey, 2009). A major player in the surveillance pathway is the exosome complex that degrades RNAs in 3' to 5' direction in the nucleus and the cytoplasm. In both compartments it is associated with various cofactors that direct it to its targets and, it is thought, thereby regulate exosome activity. For nuclear RNA surveillance the TRAMP complex is particularly important, as it marks defective RNAs with an oligo(A) tail, which serves as an activation signal for the exosome (LaCava et al., 2005b; Vanacova et al., 2005; Wyers et al., 2005). The Pol II associated Nrd1-Nab3 complex acts in transcription termination on snoRNA and cryptic RNA genes (Arigo et al., 2006b; Steinmetz et al., 2001; Thiebaut et al., 2006), which are known substrates for TRAMP/exosome processing or degradation activities, respectively. Physical interactions of Nrd1 with the nuclear exosome have been shown (Vasiljeva and Buratowski, 2006). It is therefore believed to be an important cofactor for exosome target recognition. Despite the great knowledge about the composition of the exosome complex and its various cofactors, very little is known about the identity of the RNA targets. Even more importantly, it is not known which features, within the RNA or carried by associated factors, define an exosome target.

3.2 The CRAC procedure

In order to identify new targets for the nuclear RNA surveillance machinery I applied an RNA protein crosslinking approach, CRAC (Crosslinking and analysis of cDNA) (Granneman et al., 2009) to find the RNAs that are bound by the surveillance factors Air1, Air2, Nrd1, Nab3 and Trf4. For this purpose yeast strains were constructed expressing genomically encoded C-terminal tagged Air1-HTP, Air2-HTP, Nrd1-

HTP, Nab3-HTP and Trf4-HTP (HIS-TEV-ProtA tag, Figure 3.1 A). A wild type strain further referred to as 'no tag', as well an Air2-TAP strain served as negative controls during the procedure. Rrp9-HTP, a kind gift from Sander Granneman, was used as a positive control for the crosslinking, since he had shown that this factor crosslinked very well to a single target, U3 snoRNA (Granneman et al., 2009).

The protocol is described in depth in Chapter 2.2.6 and Figure 3.1 B gives an overview of the major steps. In brief, the experiment can be summarised as follows. Cultures expressing an HTP-tagged protein, no tag or TAP control were crosslinked *in vivo* with 254 nm UV light in a Stratalinker prior to extract preparation. A two-step purification of the crosslinked RNPs was carried out. In the first step, the ProteinA tag was bound to IgG and the protein was eluted by cleavage with TEV protease. The TEV eluates were then digested with RNase A and T1 to generate a protein 'foot print' on the target RNA. A second purification step was carried out on Ni-agarose under denaturing conditions to remove associated proteins. While the protein-RNA complexes were still attached to the Ni-beads, several enzymatic reactions were performed to modify the RNA; dephosphorylation of the 5' and 3' ends, ligation of the 3' linker, radioactive labelling at the 5' end and ligation of the 5' adapter. Following elution of the RNA-protein complexes, western blot analysis was performed to monitor protein purification. For cloning of the RNA, regions of the nitrocellulose membrane containing the radioactive protein-RNA complexes were excised and the RNA was eluted by digestion of the proteins with proteinase K. The RNA was then reverse transcribed and the cDNA libraries were either cloned and Sanger sequenced or submitted for SOLEXA sequencing.

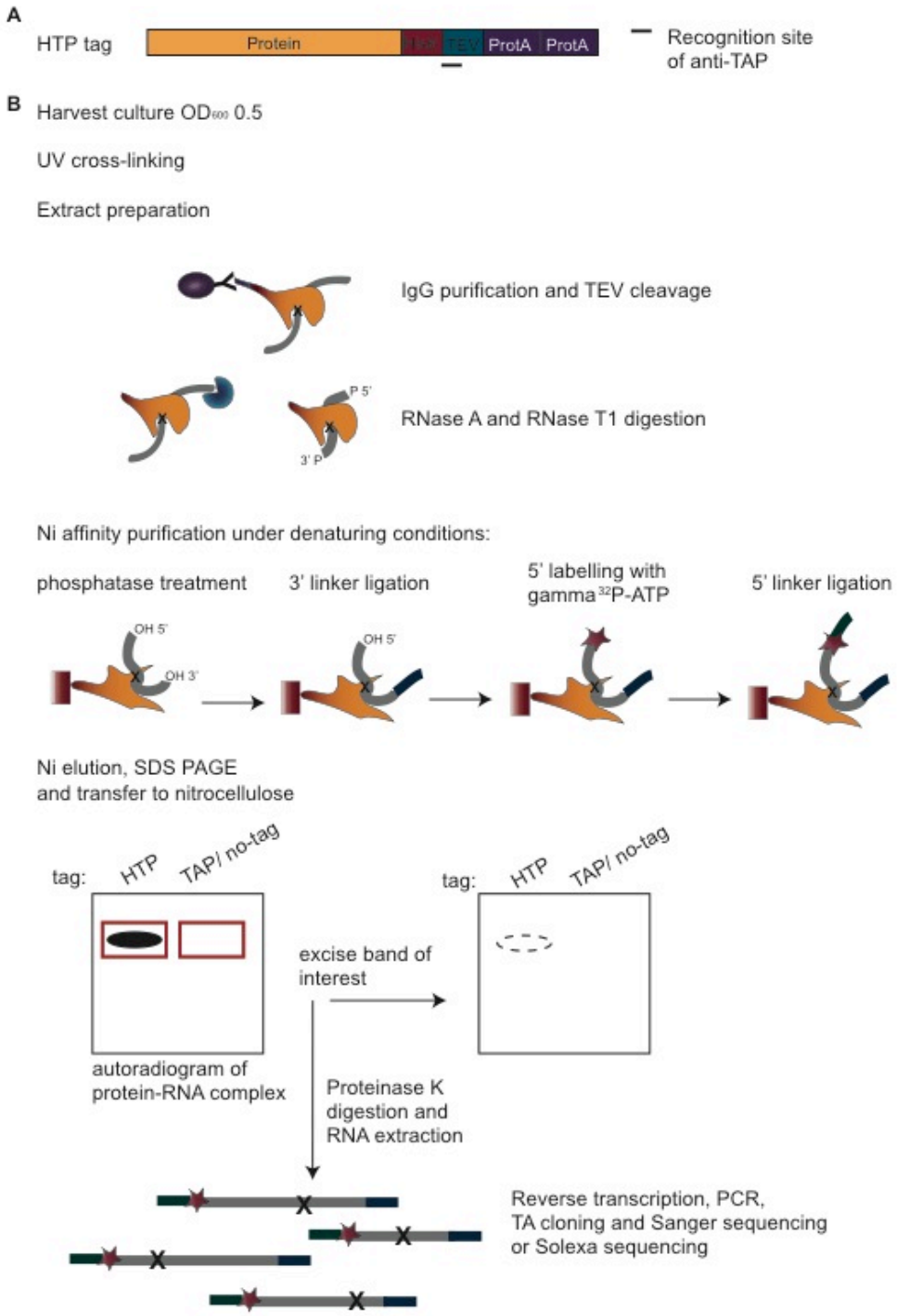

Figure 3.1 The CRAC technique
(A) Schematic representation of a protein fused to the HTP tag.
(B) Flowchart giving an overview of the experimental procedure of the CRAC technique.

3.3 Test crosslinking of Air1, Air2 and Rrp9

The procedure was first applied to the putative RNA binding proteins of the TRAMP complex Air1 and Air2 (LaCava et al., 2005a; Vanacova et al., 2005; Wyers et al., 2005). Strains were constructed expressing genomically encoded C-terminal tagged Air1-HTP and Air 2-HTP. In the experiment Air1-HTP and Air2-HTP strains were crosslinked as described above, using Rrp9-HTP (Granneman et al., 2009) as a positive control for crosslinking. Air2-TAP was used as a negative control here, since the calmodulin binding peptide of the TAP tag, exchanged for a HIS6 in the HTP tag, should not be enriched in the second purification step on Ni-agarose.

Enrichment of the proteins during the purification was monitored by western blotting (Figure 3.2 A and B). All tagged proteins were expressed to a similar extent (Figure 3.2 A; input) but after the first purification step it was apparent that Rrp9-HTP and Air2-TAP were more strongly enriched than Air1-HTP or Air2-HTP (Figure 3.2. A). After nickel-affinity purification no signal for Air2-TAP could be detected (Figure 3.2 B). Air2-HTP and Rrp9-HTP were significantly enriched in the nickel elution but Air1-HTP was hardly detectable (Figure 3.2 B). The efficiency of crosslinking was estimated from the amount of radioactively labelled RNA that purified with the protein. The RNAs associated with the proteins increase their mass and, due to their varying lengths, the crosslinked species appear as a smear above the protein band. For the radio-labelled complexes no signal could be detected for Air2-TAP other than two commonly detected contaminant bands at ~50 and 80 kDa (Figure 3.2 C; *), apparently representing radio-labelled IgG that persisted through the Ni-step. Rrp9-HTP crosslinked very well, as it purified with a lot of radio-labelled RNA (Figure 3.2 C). Air1-HTP copurified a low level of radioactively labelled RNA, in contrast to the western blot result, which showed hardly any enrichment (Figure 3.2 C). Air2-HTP purified well but crosslinked only very weakly to RNA (Figure 3.2 C). Overall, the crosslinking efficiency of Air1-HTP and Air2-HTP was very weak compared to other proteins tested. I tried to improve the purification by using different lysis conditions (data not shown), but was not able to enhance the signal for the

crosslinked RNA. Therefore, I decided not to proceed with cloning and sequencing of crosslinked RNAs with Air1-HTP and Air2-HTP.

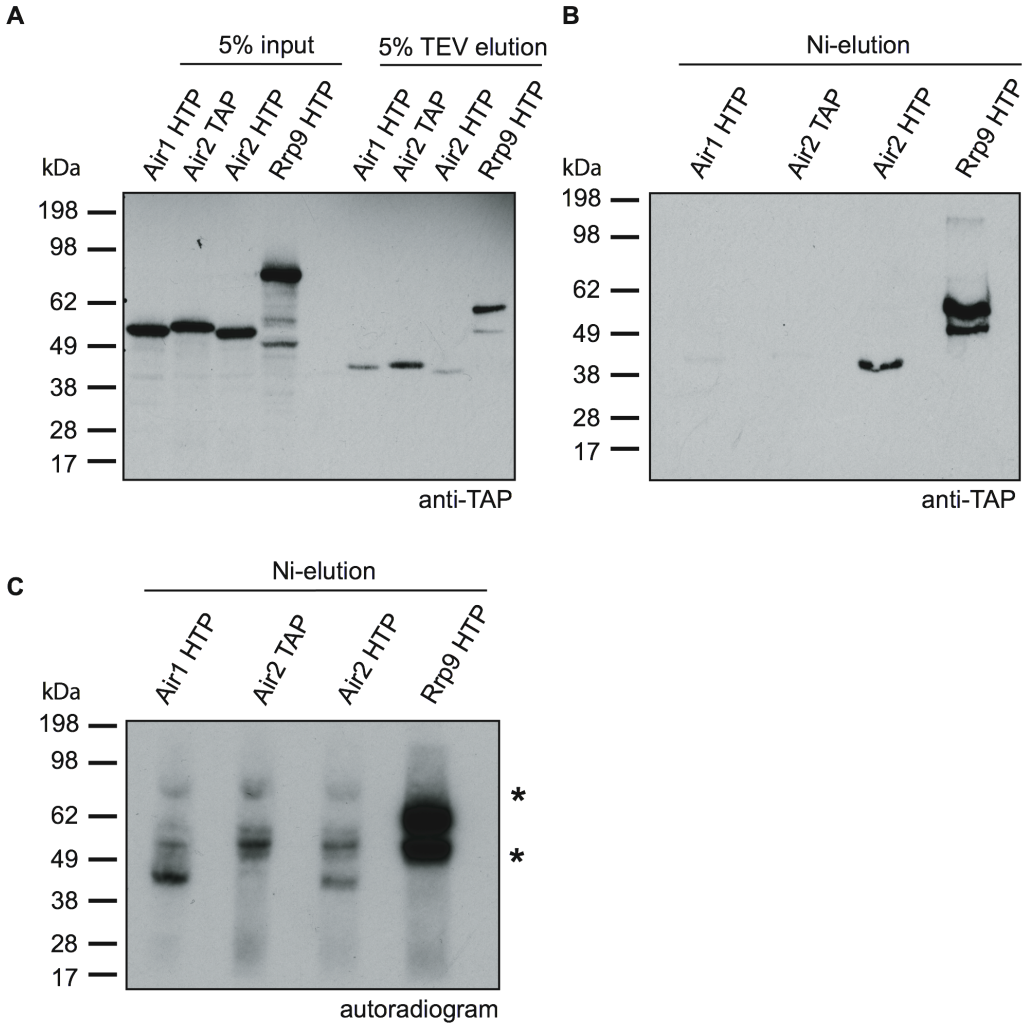

Figure 3.2 Air1, Air2 and Rrp9 test CRAC
Two step protein purification of UV crosslinked Air1-HTP, Air2-HTP, Air2-TAP and Rrp9-HTP was performed on IgG sepharose and Ni-agarose (as shown schematically in Figure 3.1).
(A) Western blot with anti-TAP antibodies of input and TEV elutions as indicated above. The migration of a protein size marker is indicated on the left.
(B) Western blot of Ni-elutions of the indicated IPs. The migration of a protein size marker is indicated on the left.
(C) Autoradiogram of radio labelled protein RNA complexes after two-step protein purification performed as in (B). Migration of common contaminants (50 and 80 kDa) is indicated by *. The migration of a protein size marker is indicated on the left.

3.4 Nrd1, Nab3 and Trf4 CRAC

In addition to the TRAMP complex, the Nrd1-Nab3 complex functions in recognition of targets for the nuclear RNA surveillance machinery (Arigo et al., 2006b; Thiebaut et al., 2006). Hence, I decided to apply the CRAC approach to the RRM containing factors Nrd1 and Nab3 as well as to the poly(A) polymerase of the TRAMP complex, Trf4.

Yeast strains were constructed expressing genomically encoded, C-terminal tagged Nrd1-HTP, Nab3-HTP and Trf4-HTP. All three strains showed wild-type growth rates (data not shown), indicating that the fusion proteins were functional. In this experiment Nrd1-HTP, Nab3-HTP, Trf4-HTP and the no tag control strain were crosslinked *in vivo* and protein-RNA complexes were purified as described above. Western blot analysis showed enrichment for all proteins in the TEV elution, whereas no signal was detected for the no tag control (Figure 3.3 A). Purification of Nrd1 protein was most efficient, but Nab3 and Trf4 also purified well (Figure 3.3 A). Crosslinking efficiency was estimated from the amount of radioactively labelled RNA that was copurified with each protein. The crosslinking efficiency of Nab3 was by far the best (Figure 3.3 B). However, Nrd1 and Trf4 both crosslinked well, compared to the Air proteins and Rrp9 (compare Figure 3.3 B 1 h exposure and Figure 3.2. C over-night exposure). The contaminant band at ~50 kDa (*) was also visible. Therefore, I continued with the cloning procedure for Nrd1, Nab3 and Trf4. The experiment was carried out in three biological replicates. The indicated regions (red squares in Figure 3.3 B) were excised from the membrane and the recovered RNAs were reverse transcribed. A small amount of the cDNA library was amplified by PCR and analysed on a 3% TBE agarose gel (Figure 3.3 C). An oligo-dimer was amplified and migrated at approximately 70 bp (marked with a *). PCR products containing the 5' and 3' linker as well as the RNA targets were excised from the indicated regions from the gel (red squares in Figure 3.3 C), gel-extracted, cloned and Sanger sequenced. For two further, independent experiments, the libraries were submitted for Solexa sequencing. In the initial Solexa sequencing analyses, samples derived from crosslinking of Nab3, Nrd1, Trf4 and the negative control were

independently sequenced. During the second round of experiments barcoded 5' linkers were used and all samples were pooled after elution from the Ni beads (Figure 3.3 B; pool CRAC) with the barcodes allowing the identification of the different input samples in the resulting sequence data.

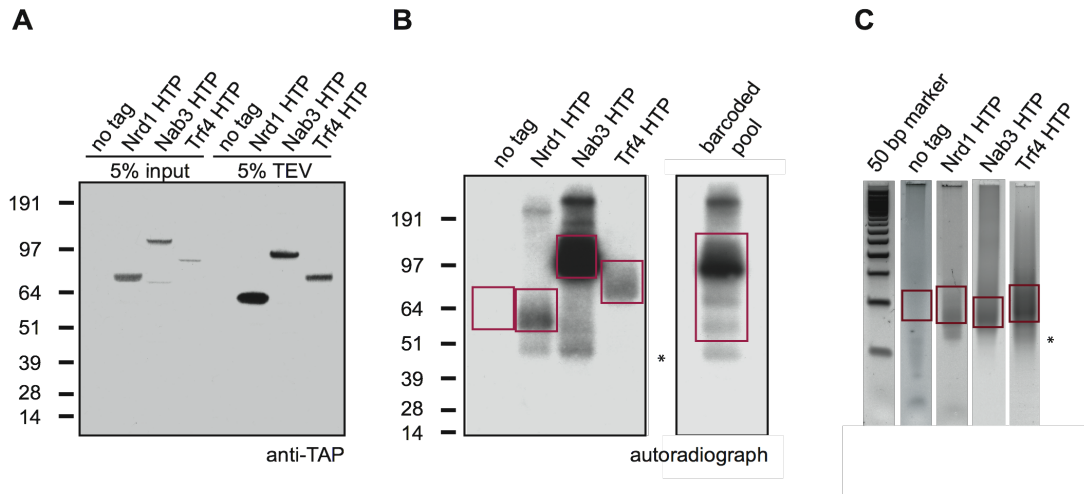

Figure 3.3 Crosslinking Nrd1, Nab3 and Trf4
Crosslinking and cDNA cloning of RNAs purified with Nrd1-HTP, Nab3-HTP, Trf4-HTP and a no-tag control after two-step protein purification on IgG sepharose and Ni-agarose (as shown schematically in Figure 3.1).
(A) Western blot with anti-TAP antibody of input and TEV elution aliquots. The migration of a protein size marker is indicated on the left.
(B) Autoradiogram of radio labelled protein RNA complexes after Ni-affinity purification. Red boxes indicate regions that were cut from the membrane and further utilized for RNA purification and cDNA synthesis. Migration of a common contaminant (50 kDa) is indicated by *. The migration of a protein size marker is indicated on the left.
(C) cDNA libraries were amplified by PCR (25-30 cycles) and products were separated on 3% TBE agarose gels. Red boxes indicate regions that were cut from the gel. Migration of a DNA size marker is indicated on the left and the migration of an oligo dimer is indicated by *.

Bioinformatics analysis was carried out in collaboration with Dr. Grzegorz Kudla as described in (Granneman et al., 2009; Wlotzka et al. 2011). In brief, Solexa sequence reads above a certain quality threshold were selected, the linker sequences were removed and the inserts mapped to the yeast genome with the NOVOALIGN algorithm (www.novocraft.com). Distribution of reads over the genome, as well as mutations or deletions within the read were displayed in the Affimetrix Integrated Genome Browser (www.affimetrix.com). Plots showing the distribution of reads

over a single gene or chromosomal region were generated using the open source software gplot. Solexa datasets were analyzed separately and results shown represent averages over both experiments, unless stated otherwise. Graphs over single genomic locations are shown for one representative experiment only. Notably, in each case the low- and high-throughput datasets were similar in their distribution of targets and presence of oligo(A) tails.

For certain parts of the analyses, datasets for the surveillance factors were compared to crosslinking data of the unrelated snoRNP factor Nop58 which is not expected to associate substantially with surveillance substrates. This experiment was performed by Dr. Sander Granneman in the lab and the unpublished data made available for comparisons.

For each of the tagged strains 2-9M sequence reads were obtained and approximately half of these could be mapped to the genome. In contrast only 23K reads could be recovered and mapped for the non-tagged control in the first dataset and 200K reads in the barcoded sample. Unfortunately, after the second, barcoded Solexa dataset had been collected we noted that in rare cases, the barcoded oligo used for the Nab3 samples could undergo a single nucleotide deletion, of unknown origin, leading to its misreading as a negative control sample. In consequence the second negative control dataset is contaminated with an unknown number of Nab3 hits. Comparison to negative control samples for single genes therefore used only the first dataset. All genome-wide analyses, such as A tail analyses, however were compared to the average of both negative control datasets. Many of the reads in the different samples could not be successfully mapped to the genome because the sequences obtained were too short to be assigned to the correct location, the quality assigned by the machine was too low or, in case of the barcoded samples, the entire barcode could not be identified.

The three tested factors each associated with several classes of RNA (Figure 3.4 A-D). For Trf4-HTP almost 50% of all sequences came from Pol I transcribed (pre-)rRNA targets, consistent with the previously reported role of the TRAMP complex

in pre-rRNA surveillance (Dez et al., 2006), and the degradation of truncated fragments generated by transcriptional pausing and R-loop formation in the 18S rRNA 5' region (El Hage et al., 2010). In contrast, Pol I transcribed (pre-)rRNAs were largely absent from the Nrd1 and Nab3 data sets despite their very high abundance in total RNA (Figure 3.4 F). The major crosslinking site for Trf4 resided in the 5' region of 18S in good agreement with the reported major site of polymerase pausing and pre-rRNA truncation (El Hage et al., 2010). The negative control data set also mostly contained rRNA targets (Figures 3.4 D and F), along with mRNA hits. The rRNA hits in the control largely corresponded to fragments in the 3' region of 25S rRNA, that were identified as common contaminants in CRAC analyses (Figure 3.4 F red box; compare (Granneman et al., 2009). mRNAs recovered in the control mainly corresponded to *HSP82*, a highly abundant chaperone, and *FMP16*, a putative protein of unknown function. Nab3 also frequently bound *HSP82* mRNA, but none of the other factors associated significantly with either of the two mRNAs.

Generally, the datasets for the heterodimeric complex proteins Nrd1 and Nab3 showed similarity concerning the different groups of RNAs that they associate with compared to Trf4 (Figure 3.4 A-C). However, genome wide or chromosome wide distribution of hits, as shown for Chromosome IV (Figure 3.4 E), did not always show strong overlap between Nrd1 and Nab3 targets. Nab3 appeared to target a larger number of different RNAs than Nrd1 (Figure 3.4 E). Statistical analyses of the total number of targeted transcripts confirmed that Nab3 targeted more different mRNAs, antisense (as)RNAs, intergenic transcripts and annotated cryptic transcripts than Nrd1 (Figure 4.1 B and C Chapter 4). This is consistent with the more efficient Nab3 crosslinking relative to Nrd1 to RNA (Figure 3.3 B). This suggests the hypothesis that Nab3 is the major RNA binding moiety of the complex while Nrd1 mainly facilitates protein-protein interactions.

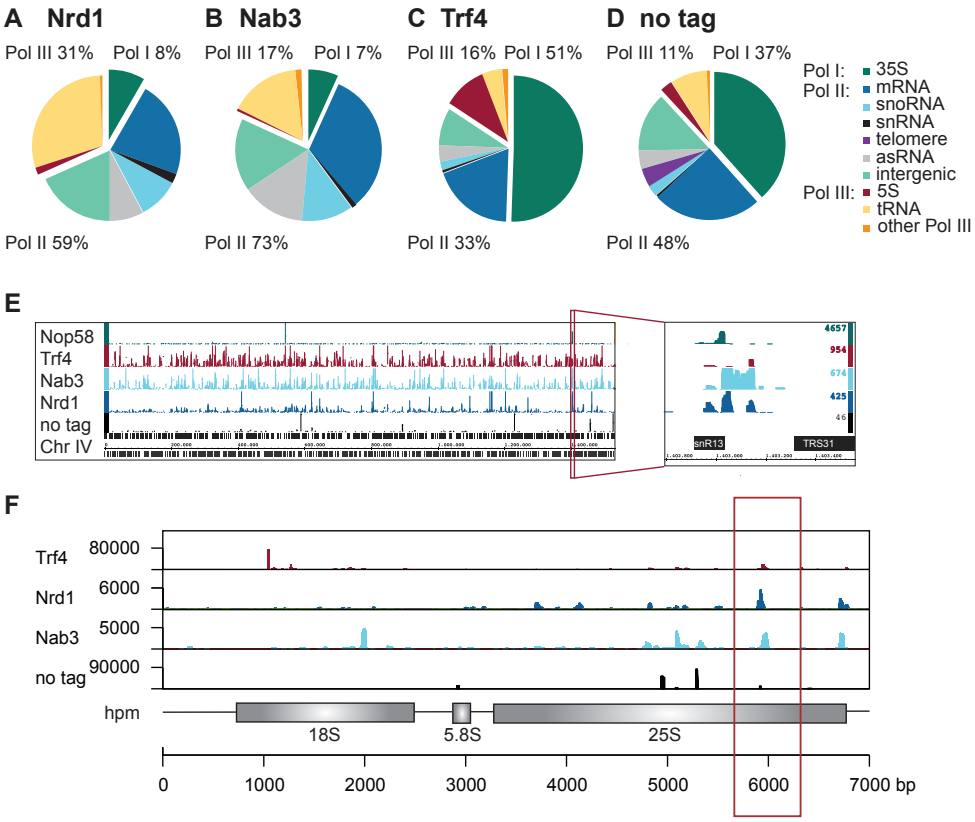

Figure 3.4 Overview of crosslinked RNA targets
(A)-(D) High-throughput sequencing of cDNA libraries generated from crosslinked RNAs associated with purified Nrd1-HTP, Nab3-HTP, Trf4-HTP and a no-tag control. Sequencing data was mapped to the yeast genome using NOVOALIGN. Pie charts illustrate the proportion of mapped reads corresponding to classes of RNAs in the indicated IP.
(E) Single track distribution of CRAC hits of the indicated IP along chromosome IV.
(F) High-throughput sequencing reads of RNAs associated with the indicated proteins are plotted over the 35S pre-rRNA. The number of hits per one million sequences is displayed.

Other stable RNAs - snRNAs and snoRNAs - were found in all datasets (Figures 3.4 A-E), consistent with reported roles for Nrd1-Nab3 in their transcription termination (Steinmetz et al., 2001) and processing or surveillance by TRAMP (Grzechnik and Kufel, 2008). Comparison of the crosslinking sites on snoRNAs for Nrd1 and Nab3 with the snoRNP factor Nop58 revealed that interaction sites are quite distinct. Nop58 preferably associated with snR13 (Figure 3.4 E) and other boxC/D snoRNAs over the internal boxD' element. This is different from the preferred surveillance

factor binding sites, demonstrating that the crosslinking method can identify specific binding sites for factors on target RNAs.

In all surveillance factor datasets a surprisingly large number of sequences were mapped to mRNAs (19-31%), suggesting either the presence of abundant cryptic sense RNAs, or nuclear pre-mRNA turnover that is more active than anticipated. The sense hits were distributed across many ORFs and are therefore unlikely to correspond to the unstable, promoter-associated transcripts that are detected in strains lacking TRAMP and exosome activities (Davis and Ares, 2006; Wyers et al., 2005).

A vast number hits from each dataset were mapped to intergenic regions (unannotated, not overlapping with any annotated feature and at least 50 nt away from any annotation) or were antisense to protein-coding genes; 26% of all reads for Nrd1, 30% for Nab3 and 13% for Trf4. A number of cryptic unstable intergenic transcripts (CUTs) were shown to be stabilized by loss of Trf4, Nrd1 or Nab3 (Arigo et al., 2006b; Thiebaut et al., 2006; Wyers et al., 2005) and the previously identified CUTs NEL025c, NGR060w and NPL040c were recovered in the CRAC analysis. Similarly, a small number of asRNAs were previously reported to be stabilized (Arigo et al., 2006b). The RNAs running antisense to the genes *RPR2* and *FMP40* were recovered in the CRAC analyses. mRNA targets as well as cryptic antisense and intergenic RNAs will be discussed in detail in Chapter 4.

The most unexpected feature in the initial analyses of the CRAC data was the apparent association of the RNA Pol II associated factors Nrd1 and Nab3 with transcripts generated by Pol III, which comprised 31% and 17% of Nrd1 and Nab3 hits, respectively. These were predominantly precursor RNAs, some of which were oligoadenylated implicating them as TRAMP targets. Trf4 was known to be involved in surveillance of defective 5S and $tRNA_i^{Met}$ (Kadaba et al., 2004; Kadaba et al., 2006; Schneider et al., 2007; Vanacova et al., 2005) and Pol III RNAs made up 16% of all mapped reads, but this was not anticipated for Nrd1-Nab3.

3.5 Identification of snoRNA targets and consensus binding motifs

To validate the crosslinking I analysed the sequencing data for the presence of known targets of Nrd1, Nab3 and Trf4. Transcription termination and processing of non-polyadenylated Pol II transcripts including snoRNAs, snRNAs, CUTs and a small number of truncated mRNAs is mediated by Nrd1-Nab3, TRAMP and the exosome (Arigo et al., 2006a; Arigo et al., 2006b; Ciais et al., 2008; Grzechnik and Kufel, 2008; Houalla et al., 2006; Steinmetz et al., 2001; Thiebaut et al., 2006; Vasiljeva and Buratowski, 2006). snoRNA genes comprised ~10% of Nrd1-Nab3 targets and 2% of all Trf4 hits (Figure 3.4 A-E). The low throughput Sanger sequencing data yielded similar abundances.

SNR13 is a well-studied example (Steinmetz et al., 2001) and Nrd1, Nab3 as well as Trf4 bound to this transcript (Figure 3.5 A) The majority of reads mapped to two terminator elements that lie downstream of the mature 3' end of the snoRNA, rather than the body of the gene (Figures 3.5 A and C). These terminator elements include the reported consensus *in vitro* binding sequences for Nrd1 (GUAA/G) and Nab (UCUU) and their mutation leads to transcription termination defects (Carroll et al., 2004; Steinmetz et al., 2001). Terminator I was bound by both Nrd1 and Nab3, with the reads directly covering the consensus binding sequences (Figure 3.5 C).

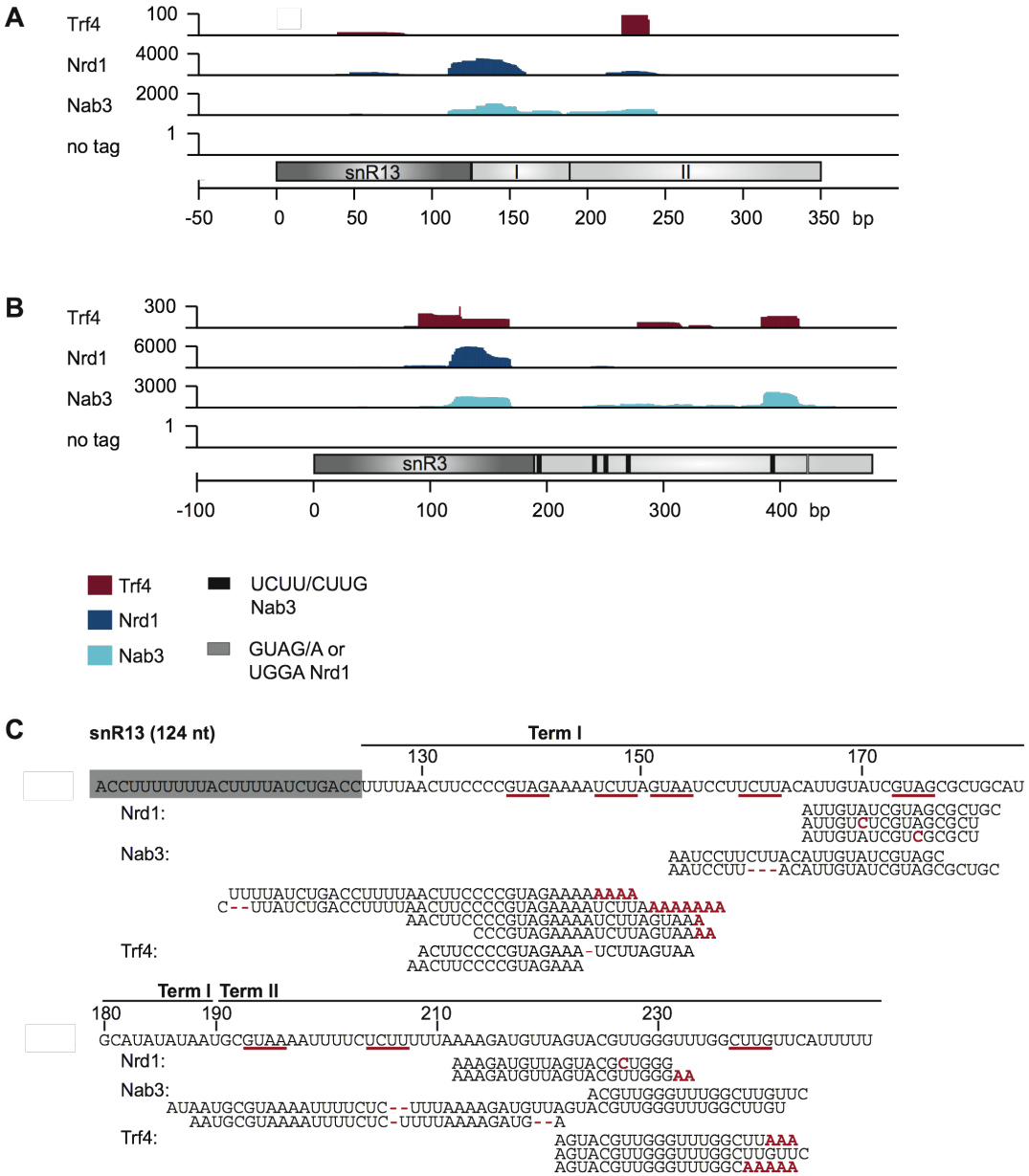

Figure 3.5 Identification of binding sites on snoRNA targets
High-throughput sequencing reads of RNAs associated with the indicated proteins are plotted over (A) the *SNR13* gene locus and downstream sequences or (B) *SNR3*. (C) Alignment of representative RNAs from the indicated IP to the terminator downstream of the *SNR13* gene. Nucleotide positions with respect to the 5' end of the *SNR13* gene and positions of the terminator elements I and II is given. The mature *SNR13* sequence is marked with a grey box. Deletions and mutations within the sequencing reads are shown in red.

Granneman et al. (2009) reported that nucleotide substitutions or deletions could indicate the precise crosslinking sites, due to the introduction of errors during reverse transcription. In order to isolate the target RNAs, the crosslinked protein is removed

by proteinase K digestion before cDNA synthesis, but at least one amino acid remains on the RNA template. They assumed that the reverse transcriptase can overcome this obstacle on the template at the site of crosslinking, but frequently introduces deletions or substitutions. These can then be used to pinpoint the exact protein binding sites.

According to this procedure I analysed mapped reads containing deleted nucleotides and was able to confidently identify the exact crosslinking sites for Nab3 in terminators I and II in each dataset. Binding sites comprised the consensus motifs at position 40 (UCUU) and 85 (UCUUUUUA; deleted nucleotides underlined; Figure 3.5 C) downstream from the end of snR13. A few Trf4 sequences in one dataset contained deletions 20 nt downstream the *SNR13* gene, positioned between Nrd1 and Nab3 binding motifs (GUAGAAAAUCUUAGUAA; Figure 3.5 C). For Nrd1 only one dataset contained deletions at +70 (GUAA) but base substitutions were frequently observed in the other dataset identifying crosslinking sites for Nrd1 approximately 50 nt downstream of *SNR13* within the consensus binding motif (GUAUCGUAG; underlined residues were substituted by cytosine; Figure 3.5 C). In addition to these point mutations, short oligo(A) tails were frequently observed in cDNAs recovered at different positions in the terminator regions. These were highly reproducible in the different datasets (Figure 3.5 C), and presumably reflect snoRNA precursor processing intermediates as observed in (Grzechnik and Kufel, 2008).

Transcription termination on *SNR3* is also impaired in *nrd1* mutants (Steinmetz et al., 2001). Nrd1 and Nab3 recovered sequences from the 3' end of the snoRNA as well as several short regions up to 300 nt downstream from the mature 3' end. These regions contained many Nab3 and fewer Nrd1 binding sites, which are bound by the respective proteins (Figure 3.5 B) suggesting that these regions contain the signals for Nrd1-Nab3 dependent snR3 termination/processing. Analysis of deletions in the sequence reads reliably revealed crosslinking of Nab3 and Nrd1 to the 3' region of snR3 in a UUG and GUGU motif as well as in the potential terminator. Nab3 crosslinking over a UCUUG motif (deleted nucleotides underlined) located 70 nucleotides downstream of the *SNR3* gene was also observed for both repeats. In

addition, recognition of the Nrd1 and Nab3 binding sites was also seen for other snoRNAs (data not shown), demonstrating the specificity of this *in vivo* crosslinking approach. These analyses therefore provide a detailed view of Nrd1-Nab3 dependent snoRNA terminator elements.

During previous studies of snoRNA transcription, consensus binding motifs for Nrd1 and Nab3 were characterized *in vitro* and *in vivo* that guide termination (Carroll et al., 2004; Steinmetz et al., 2001). It was therefore interesting to see whether these or other motifs are enriched in RNAs that were associated with Nrd1, Nab3 or Trf4.

To this end, the most abundant 4-mer sequences within the mapped reads were identified and the frequencies of their occurrence in the crosslinked data and in the genome were compared (z-score; Figure 3.6 A and B). Performing this analysis on both datasets separately yielded the same results, and one representative experiment is shown in Figure 3.6. All motif analyses were performed on genomic sequences corresponding to the reads, because raw reads frequently contained mutations in predicted protein binding sites. For Nrd1 the reported binding motifs, GUAA and GUAG, were overrepresented in the sequence data (Figure 3.6 A). However, other 4-mer sequences showed similar or greater abundance. Interestingly, UGGA, a permutated version of the previously identified motif, was the second most abundant 4-mer. No single 4-mer was present in more than 30% of reads, indicating that the presence of GUAA/G is not strictly required to recruit Nrd1 *in vivo* or that other signals in the RNA, such as RNA structure, can trigger Nrd1 binding. For both Nab3 Solexa datasets the previously identified UCUU motif was the second most abundant 4-mer in the dataset. Only a variant of this motif, CUUG scored higher and 65% of reads contained one of these motifs. Alignment of longer sequences revealed that UCUUG is indeed the core binding motif for Nab3 *in vivo* (Figure 3.6 B and C).

Analysis of the Trf4 data set did not reveal preferences for a specific motif. This is not surprising since the poly(A) polymerase should only associate with the very 3' end of the RNA for oligoadenylation and recognition of a motif would be expected to be mediated by the RNA binding proteins of the complex, Air1 or Air2. The control dataset did not contain enough mapped reads to carry out this analysis.

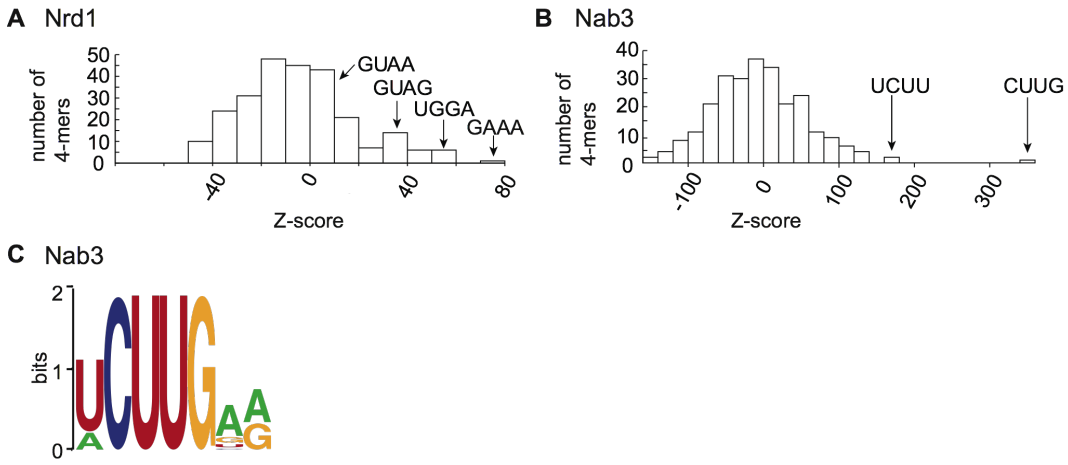

Figure 3.6 Identification of consensus binding motifs for Nrd1 and Nab3
(A) and (B) Statistical overrepresentation scores of 4-mer sequences in Nrd1 (A) and Nab3 (B).
(C) Consensus binding motif for Nab3, computed by aligning the top 50 k-mers (k= 4 through 8) with highest z-scores.

3.6 Known ncRNA targets of the nuclear RNA surveillance machinery

Cryptic RNAs transcribed by Pol II are scarcely detectable in wild type cells, due to efficient degradation by the surveillance machinery. This involves transcription termination by the Nrd-Nab complex followed by oligoadenylation by TRAMP and degradation by the nuclear exosome (reviewed by Berretta and Morillon, 2009; Jacquier, 2009). Various cryptic transcripts have been shown to be stabilized by mutants in any of the above complexes (Arigo et al., 2006b; Houseley et al., 2007a; Thiebaut et al., 2006; Wyers et al., 2005). One of these is the IGS1-R CUT, transcribed from the intergenic spacer region of the rDNA repeat, situated between the 5S gene and the 35S transcription unit (Houseley et al., 2007a; Kobayashi and Ganley, 2005). The polyadenylated transcript is stabilized by mutants of the nuclear exosome, TRAMP and Nrd-Nab complexes. The IGS1-R ncRNA is proposed to help maintain stability of the rDNA repeat (Houseley et al., 2007a; Vasiljeva et al., 2008b).

The IGS1-R ncRNA was frequently recovered in the crosslinking experiments with all three factors (Figure 3.7). Analysis of deletions within the reads revealed binding

sites for Nab3 lying 18 nt (CUCUCU) and 125 nt (UCUUACAUCUUUCUU, deleted nucleotides underlined) downstream of the TSS of the IGS1-R transcript. This was, however, observed in only one experiment; in the second dataset as well as in the Nrd1 and Trf4 datasets none of the sequences contained deletions.

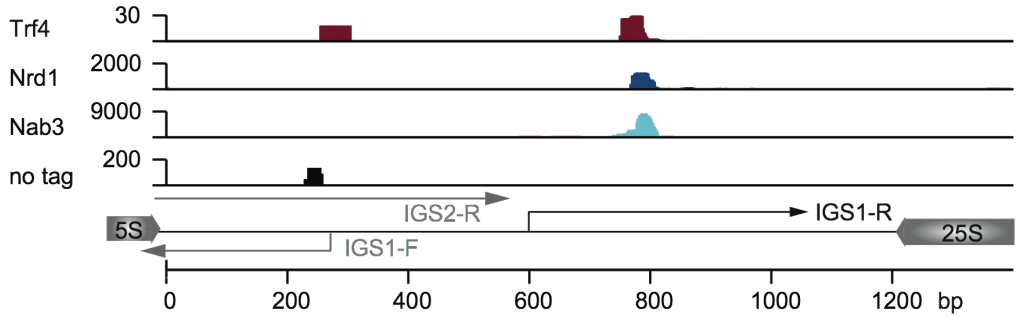

Figure 3.7 Identification of crosslinks to cryptic Pol II transcripts in the intergenic spacer of the rDNA repeat
High-throughput sequencing reads of RNAs associated with the indicated proteins are plotted over the intergenic spacer region of the rDNA repeat (plus strand). A schematic representation of the spacer region with the flanking rRNA genes (Pol I and III) and ncRNA transcripts (Pol II) is displayed below.

SRG1 is an ncRNA that partially overlaps with the promoter of the downstream gene *SER3*. The *SER3* gene encodes for an enzyme required for serine biosynthesis and the expression of its mRNA is repressed in rich media (Martens et al., 2004; Martens et al., 2005). Under these growth conditions, transcription of the upstream *SRG1* gene can be detected and has been shown to repress *SER3* mRNA transcription by a transcription interference mechanism (Martens et al., 2004; Martens et al., 2005). Furthermore, it is known that *SRG1* is an oligoadenylated transcript that is degraded in the nucleus in an Nrd-Nab/TRAMP/exosome dependent manner or in the cytoplasm by the 5' exonuclease Xrn1 after decapping (Arigo et al., 2006b; Thiebaut et al., 2006; Thompson and Parker, 2007).

Nrd1, Nab3 and Trf4 were found associated with *SRG1* ncRNA (Figure 3.8), mainly over the middle of the RNA. No hits could be found for the downstream gene *SER3* with Trf4, and only very few with Nrd1 and Nab3 (1080 nt downstream of the *SRG1* TSS in the middle of the *SER3* ORF).

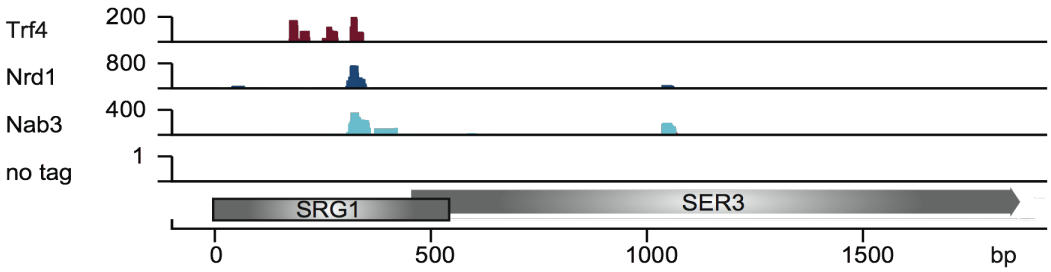

Figure 3.8 Identification of binding sites to *SRG1* ncRNA
High-throughput sequencing reads of RNAs associated with the indicated proteins are plotted over *SRG1* ncRNA and the downstream gene *SER3*. A schematic representation of the genomic locus is displayed below.

3.7 Known mRNA targets of the Nrd1-Nab3 transcription termination pathway

Transcription termination of Pol II on mRNA genes usually occurs after the primary transcript has been processed by the cleavage and polyadenylation machinery. In the 'torpedo model' the 5' exonuclease Rat1 chases after the still elongating RNA Pol II, degrading the downstream cleavage product and forcing Pol II to terminate (Kim et al., 2004). In contrast to this, transcription termination of non-polyadenylated Pol II transcripts, such as sn(o)RNAs and CUTs, is induced by a mechanism involving the Nrd1-Nab3 complex, although degradation of the nascent transcript by Rat1 probably also contributes(Kim et al., 2006).

3' end formation on two mRNAs was previously shown to utilize a closely related pathway (Arigo et al., 2006a; Ciais et al., 2008; Houalla et al., 2006; Steinmetz et al., 2001) and these mRNAs were recovered in the CRAC approach. The *CTH2* mRNA is generated by post-transcriptional processing from a precursor that is 3' extended by ~1.6kb (Ciais et al., 2008). Maturation involves recognition of the pre-mRNA by Nrd1-Nab1 and subsequent 3' processing by TRAMP and the exosome. Sequences associated with Nrd1, Nab3 and Trf4, were consistent with binding to the 3' extended pre-*CTH2* RNA (Figure 3.9 A). Trf4 mainly recovered a region 800 nt downstream of the mRNA stop codon. For Nrd1 and Nab3, crosslinked sequences lay over the previously predicted cluster of binding sites located at +900 relative to

the 3' end of the ORF, supporting both the previous conclusions concerning the processing pathway and the reliability of the CRAC technique. However, no crosslinks were observed in a $(GU_3)_5$ repeat around +240 downstream of the ORF, which was identified as the major polyadenylation site of the *CTH2* mRNA.

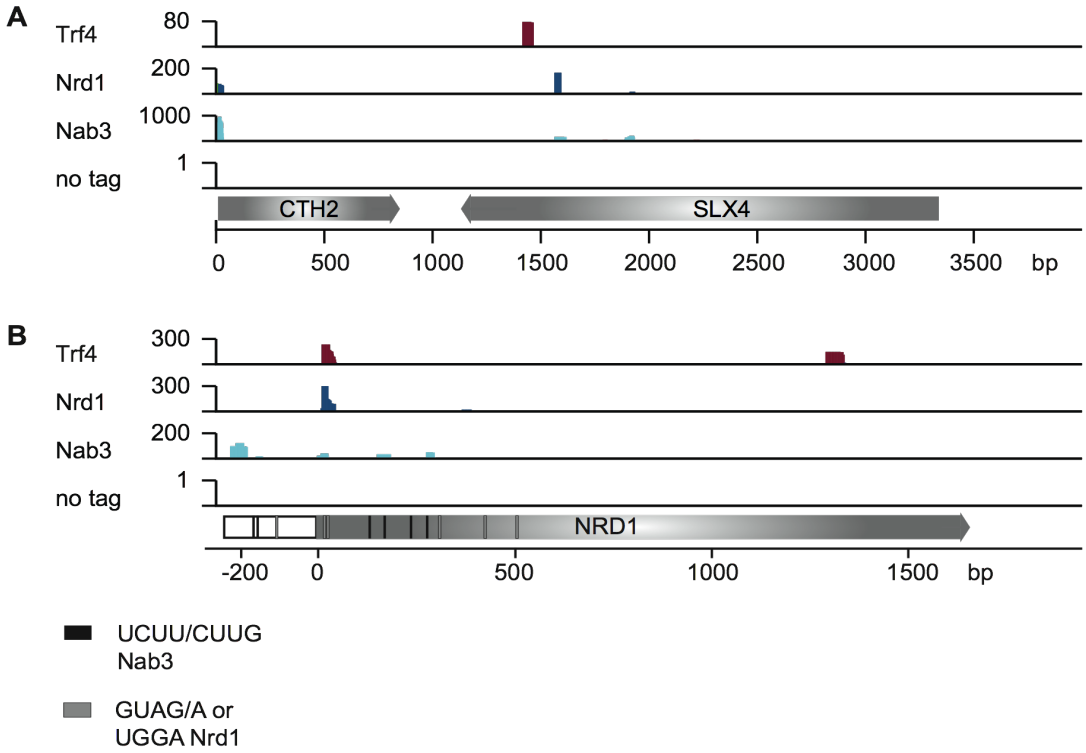

Figure 3.9 Identification of crosslinking sites on *NRD1* and *CTH2* mRNAs
High-throughput sequencing reads of RNAs associated with the indicated proteins are plotted over (A) *CTH2* genomic locus (plus strand) and (B) *NRD1* mRNA.

The 5'UTR and the 5' coding region of the *NRD1* mRNA contain several Nrd1 and Nab3 consensus binding sites, which are involved in auto-regulation of Nrd1 levels, as they direct Nrd1-Nab3 dependent, premature transcription termination (Arigo et al., 2006a; Steinmetz et al., 2001). Regulation of *NRD1* mRNA is also dependent on nuclear exosome cofactors, since loss of Rrp47 or Trf4, but not of Rrp6 leads to increased levels of full-length *NRD1* mRNA. Similar phenotypes have been observed in *nrd1* and *nab3* mutants (Arigo et al., 2006a; Houalla et al., 2006). Nrd1 and Nab3 recovered various sequences in the 5' UTR and the 5' end of the *NRD1* mRNA, including the consensus motifs (Figure 3.9 B). Distinct crosslinking sites were

identified by deletions in sequences bound to Nab3, approximately 230 nt upstream of the start codon and 100 nt into the gene.

3.8 CRAC targets are polyadenylated

Oligoadenylation of RNAs by the TRAMP complex is an important signal for degradation mediated by the exosome (LaCava et al., 2005a; Vanacova et al., 2005; Wyers et al., 2005). cDNAs associated with Nrd1, Nab3 and Trf4 were compared to the genomic sequence and analyzed for the presence of non-encoded, 3' adenosine residues. Since the RNA was fragmented with RNases A and T1, which do not cut adjacent to As, 3' oligo(A) tails should remain intact. As a control RNAs crosslinked to the snoRNP protein Nop58 were analyzed (Granneman et al., 2009), because this factor is not expected to be substantially associated with surveillance substrates.

Nrd1, Nab3 and Trf4 were each preferentially associated with RNAs carrying non-encoded oligo(A) tails (Figure 3.10 A). For Nrd1 18% of all mapped reads were oligoadenylated, whereas 6% of Nab3 and 15% of all Trf4 associated sequences carried an oligo(A) tail. Sanger sequencing analyses revealed 10-20% oligoadenylated sequences, the majority of which were tRNAs. In contrast, Nop58 and the non-tagged control recovered less than 1% and 3% of oligoadenylated sequences, respectively. This provides strong support for the recovery of *bona fide* targets for the surveillance machinery.

The A tails recovered on the CRAC targets were generally short (2-4 nt); 50% of all A tails on the were $\leq$ 2 nt for Nab3 and $\leq$ 3 nt for Nrd1 and Trf4 (Figure 3.10 B). Length distributions in the Nrd1 and Trf4 datasets showed a clear peak at A_4, with an additional peak at $A_{8\text{-}9}$ (Figure 3.10 C).

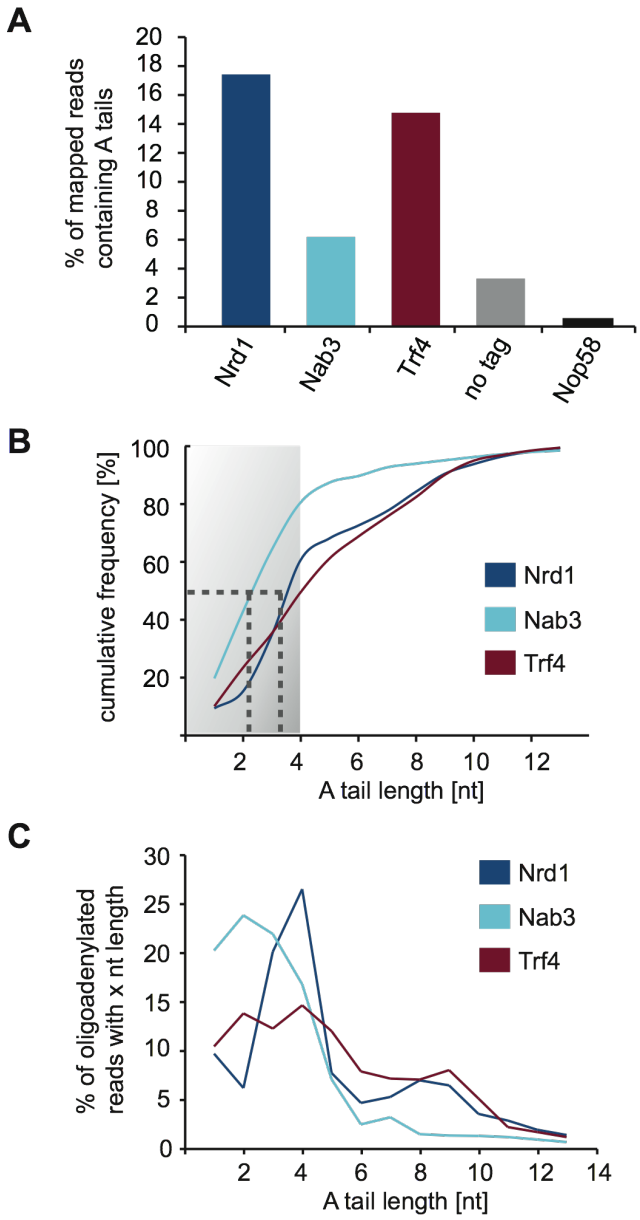

Figure 3.10 Nrd1, Nab3 and Trf4 target RNAs are oligoadenylated
(A) Bar diagram representing percentage of all mapped sequences carrying at least three non-templated terminal A residues in the indicated IP.
(B) Cumulative frequencies of non-templated oligo(A) tails on RNAs associated with the indicated protein.
(C) Length distribution of non-templated oligo(A) tails on RNAs associated with the indicated protein.

3.9 Discussion

At the time of this work, the CRAC approach had been applied only to snoRNPs and ribosomal proteins (Bohnsack et al., 2009; Granneman et al., 2009), which were expected to bind only a few, very abundant RNAs or specific motifs within the RNA. I successfully applied this approach to the RNA binding proteins Nrd1 and Nab3 and the poly(A) polymerase Trf4 (Figure 3.3). These factors each bind a large variety of RNAs, many of which have very low abundances. Overall, the components of the surveillance machinery crosslinked to a wide range of substrates (Figure 3.4), as expected, whereas the no-tag control recovered notably less RNAs (1% of the number of mapped reads in the Nrd1 dataset). In negative control datasets the majority of sequences mapped to sequences derived from the 3' end of the 25S gene (Figure 3.4), which are commonly found as contaminants (Granneman et al., 2009), or to mRNAs that are not considerably represented in the Nrd1, Nab3 or Trf4 datasets, demonstrating the reliability of the crosslinking data.

The reliability of the crosslinking data was confirmed by the recovery of many known targets (Figures 3.5, 3.7, 3.8 and 3.9) with strong enrichment for binding motifs previously identified for Nrd1 (GUAG/A) and Nab3 (UCUU) (Figure 3.5). Bioinformatics analysis was conducted to compare the abundance of the previously identified *in vitro* binding motifs within all mapped reads and the yeast genome. The motif analysis revealed indeed a strong preference of Nab3 for the known binding motif (Figure 3.6). In addition, the *in vitro* binding motif for Nab3 was amended according to the motif analysis, since UCUUG was identified as the core recognition motif *in vivo*. In the Nrd1 dataset, however, the abundance of other motifs was similar or greater that the reported motif, indicating that GUAG/A does not represent the exclusive binding site for Nrd1 *in vivo*. Given that GUAG/A were identified as elements guiding transcription termination of non-polyadenylated Pol II transcripts (Carroll et al., 2007; Carroll et al., 2004; Steinmetz et al., 2001), it is conceivable that other target groups are recognised by a different motif or a sequence-independent mechanism. Nab3 showed stronger RNA crosslinking than Nrd1 (Figure 3.3), perhaps reflecting closer RNA binding *in vivo*. It is possible that Nab3 is responsible

for the initial RNA contact, which appears to be sequence specific, followed by Nrd1 binding.

Sequence analysis of reads mapping to the terminator elements of *SNR13* identified precise crosslinking sites for Nrd1 and Nab3 over their specific binding motifs, as well as oligoadenylated processing intermediates of the snoRNA (Grzechnik and Kufel, 2008). Oligoadenylated pre-snR13 species were stabilised in strains with reduced nuclear exosome activity (*rrp6*Δ). Reported major polyadenylation sites in terminator I match the sites of oligoadenylation identified in the Nab3 CRAC data (GAAAA(A_N)UCUU(A_N)AGU(A_N), observed oligoadenylation sites in red Grzechnik and Kufel, 2008); compare to oligoadenylated sequences of Nab3 CRAC Figure 3.5 C). Two polyadenylation sites were previously identified in the terminator II element, the first one situated 85 nt downstream from *SNR13* (UUUUCUCUUUU(A_N), observed oligoadenylation site in red Grzechnik and Kufel, 2008). CRAC reads in this region the did not carry non-templated oligo(A) tails, but Nab3 was crosslinked to the UCUU motif. The second site in terminator II is located 110 nt downstream of *SNR13* (CUUGU(A_N), observed oligoadenylation site in red Grzechnik and Kufel, 2008) and Trf4 CRAC sequences were frequently oligoadenylated in that region (Figure 3.5 C); oligoadenylation sites were reliably recovered in both repeats of the experiments. The CRAC approach can therefore also identify processing intermediates. I conclude that the CRAC technique can be applied to identify unknown targets and their processing intermediates with the same accuracy.

An outstanding question was how A tail addition by TRAMP could target aberrant RNAs for degradation, while mRNAs were stabilized by polyadenylation? Bioinformatics analysis revealed that RNAs associated with surveillance factors significantly more frequently carried oligo(A) tails than RNAs crosslinked to the snoRNP protein Nop58 or the no-tag control (Figure 3.10 A), identifying them as genuine surveillance targets. Moreover, these oligo(A) tails showed a median length of approximately A_{2-4} with a smaller peak at A_{8-9} (Figure 3.10 B and C). These data show that oligo(A) tails added by TRAMP are predominately too short to bind the canonical poly(A) binding protein Pab1, which stabilizes mRNAs and stimulates

translation but requires ~A_{12} to bind (Sachs et al., 1987). Hence, oligo(A) tails on the surveillance targets are predicted to remain unprotected, providing an unstructured entry site for exonucleases including the exosome complex.

Notably, this result further implies that the RNAs identified here as surveillance substrates would predominately be overlooked in previous microarray analyses that involve oligo(dT) selection or priming for cDNA synthesis.

Chapter 4

Identification of ncRNA and mRNA targets for the nuclear RNA surveillance machinery

4.1 Introduction

Recently it became apparent that most of the genome in yeast and human cells is actively transcribed by RNA Pol II, giving rise to ncRNAs such as cryptic transcripts and antisense RNAs (asRNAs) (Neil et al., 2009; Wyers et al., 2005; Xu et al., 2009). Depending on their expression levels in the presence or absence of the nuclear exosome component Rrp6, Xu *et al.* separated the newly-identified non-protein coding Pol II transcripts into SUTs (stable unannotated transcripts), which are unaffected by loss of Rrp6, or CUTs (cryptic unstable transcripts), which are more abundant in the absence of Rrp6.

Two yeast asRNAs have been functionally analyzed and shown to participate in regulating the expression of the corresponding, sense mRNA (*GAL10as, PHO84as*) (Berretta et al., 2008; Camblong et al., 2007; Houseley et al., 2008). Repression of mRNA expression by a ncRNA may be mediated by RNA-depended gene-silencing pathways involving changes in the chromatin state (Berretta et al., 2008; Camblong et al., 2007; Houseley et al., 2008). A distinct mechanism is transcriptional interference, in which the ncRNA is transcribed across the mRNA promoter interfering with the assembly of pre-initiation complexes. Such transcription interference mechanisms were shown to control expression of *SER3* and *IME4* mRNAs (Hongay et al., 2006; Martens et al., 2004).

4.2 A large number of protein coding and cryptic Pol II transcripts are selectively targeted for RNA surveillance by Nrd1-Nab3

The CRAC approach identified many Pol II transcripts that were mapped to mRNAs, asRNAs, CUTs, SUTs and other intergenic regions (Figures 3.4 and 4.1). To assess the total number of targets, RNAs that were identified in both datasets with clusters of two or more overlapping, non-identical hits were counted. Hit clusters were identified for a total 1384 different mRNA sense strands that were not previously characterized as targets of Nrd1-Nab3 termination or nuclear surveillance and for 642 putative asRNAs (Figure 4.1 A). In addition, hit clusters were identified over

178 other intergenic regions. The identification of these ncRNA confirms that they are actively transcribed in wild-type cells, and are not only induced by mutation of the surveillance machinery.

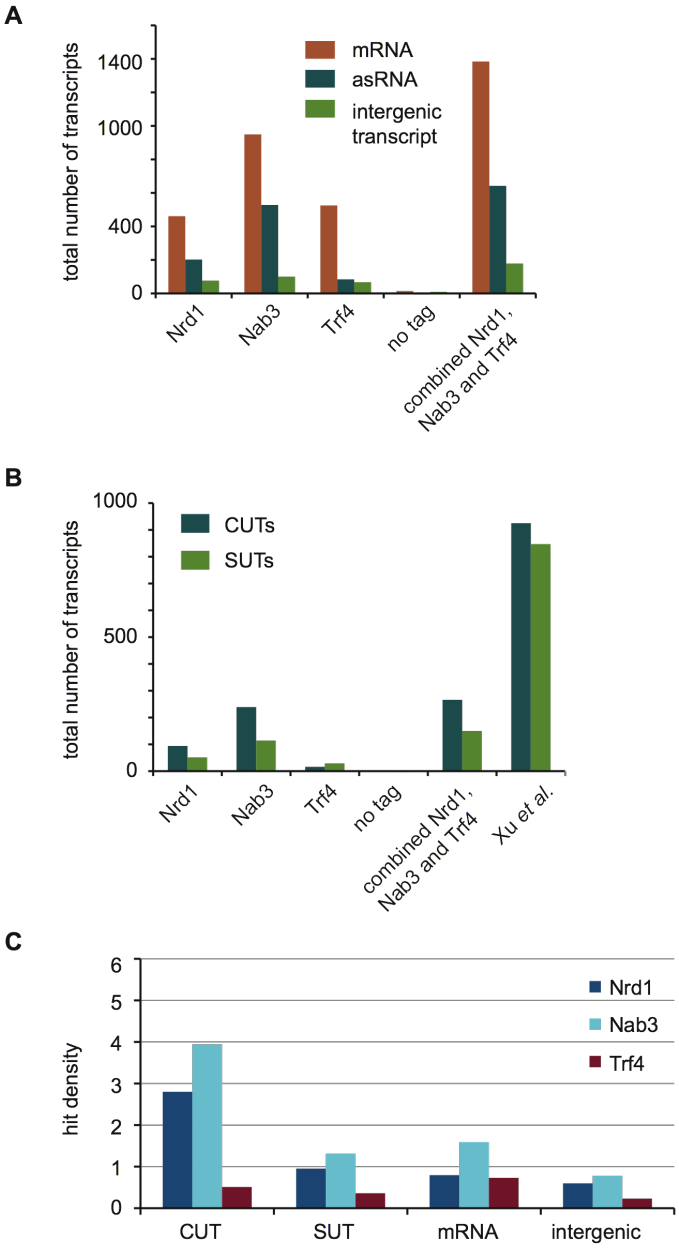

Figure 4.1 Protein-coding and cryptic Pol II transcripts are targets for the nuclear RNA surveillance machinery
(A) Clusters of high-throughput sequencing reads of RNAs were mapped to mRNAs, asRNAs or intergenic regions. The total number of different transcripts is displayed for each protein and the Nrd1-Nab3 and TRAMP complex combined. Only RNAs that were present in both datasets were counted.

(B) Clusters of high-throughput sequencing reads of RNAs were compared to the population of CUTs and SUTs identified by (Xu et al., 2009). The total number of different CUT and SUT transcripts is displayed for each protein and the Nrd1-Nab3 and TRAMP complex combined. Only RNAs that were present in both datasets were counted.
(C) Density of high-throughput sequencing reads mapped to CUTs, SUTs, mRNAs and intergenic regions, according to (Xu et al., 2009).

Comparison of the CRAC datasets with the CUTs and SUTs (Xu et al., 2009) revealed considerable overlap, with CRAC hit clusters found reliably in both datasets on 266 CUTs (29% of total) and 150 SUTs (18% of total) (Figure 4.1 B). Notably the averaged density of Nrd1 and Nab3 hits over all annotated CUTs was substantially higher than over annotated SUTs, ORFs or intergenic regions (Figure 4.1 C; densities for one experiment are shown, the second repeat yielded the same result). This strongly supports the hypothesis that Nrd1-Nab3 binding constitutes a general feature that targets ncRNAs to the exosome.

The identified RNAs are putative targets for the nuclear surveillance machinery, as well as potential regulators of gene expression. I therefore characterised selected, potential asRNAs in strains carrying mutations in the surveillance machinery.

4.3 *GAL10as* RNA

In the *GAL* cluster, a 4.0 kb long ncRNA (*GAL10as*) is expressed when transcription of the *GAL* mRNAs is repressed in glucose media (Houseley et al., 2008). The asRNA transcript initiates within the *GAL10* gene, runs antisense through *GAL10*, across the *GAL1-10* bidirectional promoter and sense through *GAL1*. The *GAL10as* is subject to TRAMP-dependent degradation and present at only 0.07 copies per cell (i.e. about one cell in 13 has a copy of the RNA at steady state). It was shown that *GAL10as* is able to direct chromatin modifications over the *GAL* cluster, helping to maintain the repressive state in glucose media.

Despite its very low abundance, association of Trf4 with the *GAL10as* could be observed in the CRAC experiment in a wild type background (Figure 4.2). The CRAC data also revealed crosslinking of Nrd1 and Nab3 to *GAL10as*. Depletion of

Nrd1 or Nab3 increased the level of the *GAL10as*, confirming that it is indeed a target (data not shown). The recovery of this very rare transcript demonstrates the sensitivity of the CRAC technique in identifying low abundance targets of the surveillance machinery.

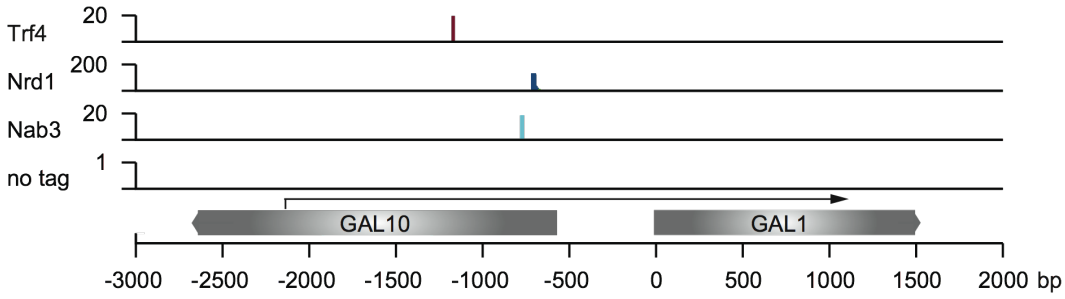

Figure 4.2 *GAL10as* RNA
High-throughput sequencing reads of RNAs associated with the indicated proteins are plotted over the *GAL10-GAL1* locus. Sequencing reads are shown for the plus strand, encoding the *GAL10as* RNA as well as *GAL1* mRNA.

4.4 Characterization of unknown asRNA transcripts

To validate the CRAC data, the expression of selected candidate asRNAs was examined in surveillance mutants. The most abundant asRNA target from the Trf4 crosslinking data was *HPF1as*, which was recovered as frequently as some snoRNAs (Figure 4.4 A). Trimethylation of histone 3 lysine 4 (H3K4) serves as a mark for active RNA Pol II transcription. Comparison of the genome wide H3K4 trimetylation data (Kirmizis et al., 2007) with the location of the potential *HPF1as* RNA, showed a large peak of H3K4 trimethylation at the 3' end of the corresponding mRNA, supporting the existence of a transcript initiating in that region and running antisense to the *HPF1* mRNA. Other potential asRNAs were identified in the same way and association of Nrd1 and Nab3 was also observed, but with fewer hits than for *HPF1*.

Transcription start sites (TSS) of potential asRNAs were mapped for two candidate asRNAs, *CAF17as* and *HPF1as,* by random primed 5' RACE in *trf4*Δ and WT strains (Figure 4.3 A). The TSS for *HPF1as* was mapped 265 nt downstream of the

mRNA stop codon (Figure 4.3 B). For *CAF17as* two TSSs were identified one lying 10 nt upstream and one 10 nt downstream of the stop codon of the corresponding mRNA (figure 4.3 C).

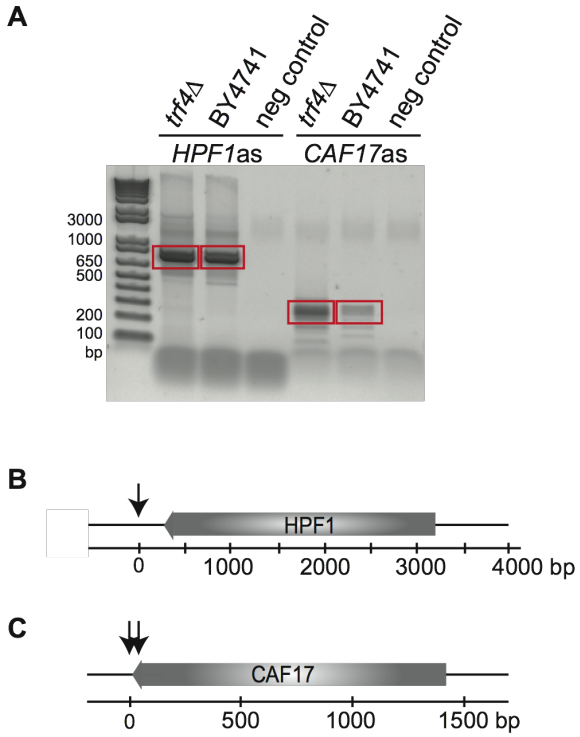

Figure 4.3 Mapping of asRNA transcription start sites by 5' RACE
(A) 5' RACE PCR was performed on RNA from wild-type and *trf4Δ* cells, with primers located 494 bp from the 3' end of the annotated *HPF1* ORF and 256 bp from the 3' end of the annotated *CAF17* ORF. DNA products were stained with SYBR Safe. The indicated band was excised, cloned and sequenced.
(B) and (C) Schematic representation of the chromosomal loci encoding *HPF1* (B) and *CAF17* (C). Transcription start sites of the asRNAs are marked with an arrow. The scale on the bottom displays bp with respect to the transcription start site of the asRNA.

Association of the asRNAs with Trf4, Nrd1 and Nab3 suggested that these factors are involved in the surveillance of the antisense transcripts. Since these RNAs are usually rapidly degraded, I investigated to what extent mutations in the surveillance machinery stabilize the identified asRNAs.

To this end the genes of the essential factors Nrd1 and Nab3 were placed under control of a conditional promoter. Genomically encoded N-terminal HA tagged *P_{GAL}::NRD1*, *P_{GAL}::NAB3* and *P_{GAL}::NRD1 P_{GAL}::NAB3* strains were constructed (Lafontaine and Tollervey, 1996) and the proteins were depleted in glucose media (Figure 4.4). Growth was monitored over time and cells were harvested for RNA preparations when the growth rate of the mutant was significantly reduced compared to the isogenic wild type strain (BY4741).

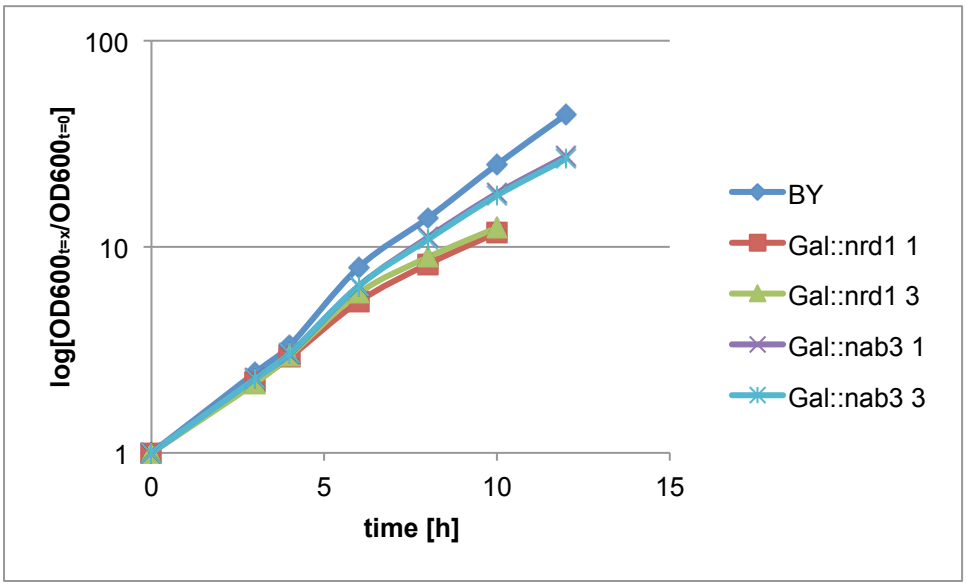

Figure 4.4 Growth curves during Nrd1 and Nab3 depletion
Growth curves of the indicated strains expressing genomically encoded, N-terminal HA-tagged Nrd1 or Nab3 under control of the *GAL* promoter. BY4741 (BY) is the isogenic WT. Strains were pre-grown in YPGalSuc and then shifted to YPD for the indicated times. The growth of two individual clones for each mutant is displayed.

Northern hybridisation was carried out in strains depleted for Nrd1 and Nab3 or in mutants of the TRAMP and exosome complexes (*trf4*Δ, *trf5*Δ and *rrp6*Δ, purchased from Euroscarf). Strong asRNA accumulation was observed in strains depleted for Nrd1 or Nab3, or lacking either Rrp6 or Trf4, but not in strains lacking the homologous poly(A) polymerase Trf5 (Figure 4.5 B *HPF1as* and 4.6 B *CAF17as*). All asRNAs detected were long (0.5 – 8kb) but notably heterogeneous in size, with multiple bands being visible by northern hybridization. Heterogeneity was also observed for previously analyzed asRNAs and intergenic RNAs (Arigo et al., 2006b; Thiebaut et al., 2006), and may be a common feature of yeast ncRNAs.

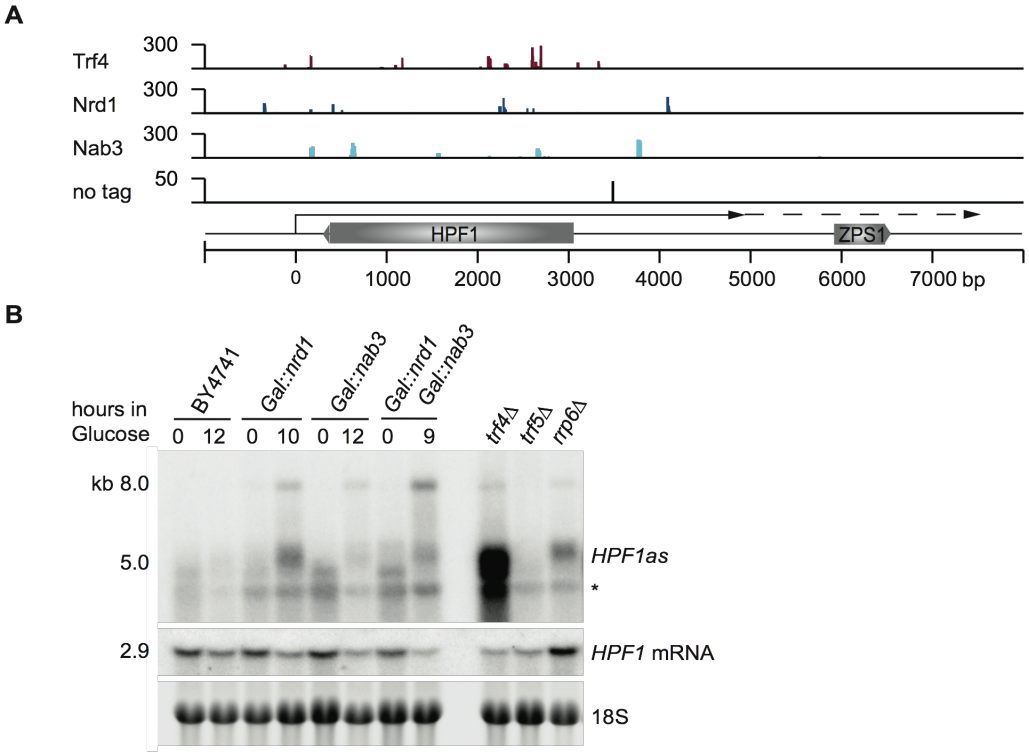

Figure 4.5 Characterisation of *HPF1as* RNA
(A) High-throughput sequencing reads of RNAs associated with the indicated proteins are plotted over the *HPF1as* RNA. An arrow represents the major asRNA species; the detected read-through product (as observed in B) is illustrated as a dashed arrow.
(B) Northern blot of total RNA prepared from the indicated strains in exponential growth. P_{GAL}*::NRD1*, P_{GAL}*::NAB3* and P_{GAL}*::NRD1* P_{GAL}*::NAB3* and BY4741 strains were pre-grown in YPGalSuc and then shifted to YPD for the indicated times. *trf4*Δ, *trf5*Δ and *rrp6*Δ were grown in YPD. RNAs were separated on denaturing glyoxal BPTE agarose gels (1%) and transferred to nylon. Riboprobes were directed against the indicated species and the size of the detected RNAs is indicated. Ethidium bromide staining of 18S rRNA served as loading control.

For both *CAF1as* and *HPF1as*, the major bands seen in *trf4*Δ and *rrp6*Δ strains were shorter than in strains depleted for Nrd1 or Nab3. This would be consistent with a role for Nrd1 and Nab3 in transcription termination on these asRNAs, with the longer RNAs representing read-through products. The major form of *HPF1as* (5.0 kb) was also weakly expressed in the isogenic WT strain (Figure 4.5 B). There was, however, no clear correlation between asRNA accumulation and expression of the sense mRNA.

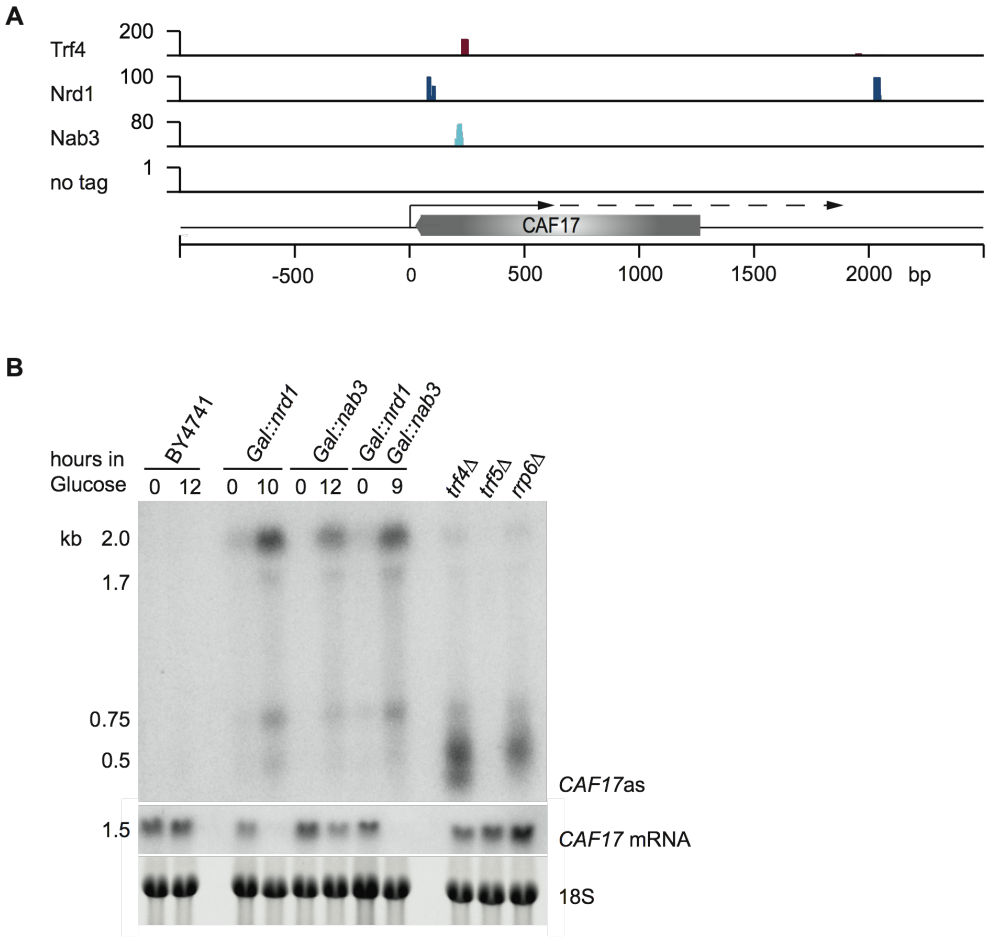

Figure 4.6 Characterization of *CAF17as* RNA
(A) High-throughput sequencing reads of RNAs associated with the indicated proteins are plotted over the *CAF17as* RNA. An arrow represents the major asRNA species; detected read-through product (as observed in B) is illustrated as a dashed arrow.
(B) Northern blot of total RNA prepared from the indicated strains in exponential growth. P_{GAL}*::NRD1*, P_{GAL}*::NAB3* and P_{GAL}*::NRD1* P_{GAL}*::NAB3* and BY4741 strains were pre-grown in YPGalSuc and then shifted to YPD for the indicated times. *trf4*Δ, *trf5*Δ and *rrp6*Δ were grown in YPD. RNAs were separated on denaturing glyoxal BPTE agarose gels (1%) and transferred to nylon. Riboprobes were directed against the indicated species and the size of the detected RNAs is given. Ethidium bromide staining of 18S rRNA served as loading control.

4.5 mRNA targets of Trf4, Nrd1 and Nab3

mRNA turnover predominantly takes place in the cytoplasm and is initiated by deadenylation, followed by decapping and 5′ to 3′ degradation by the cytoplasmic exonuclease Xrn1. Although 5′ to 3′ degradation by Xrn1p appears to be the common

mode of decay at least in yeast, mRNAs can also be degraded from the 3′ end by the cytoplasmic exosome (reviewed by Parker and Song, 2004). A few examples have been reported in which specific groups of mRNAs are turned over in the nucleus under certain (growth) conditions (Bousquet-Antonelli et al., 2000; Kuai et al., 2005; Lee et al., 2005; Reis and Campbell, 2007). mRNAs encoding proteins involved in iron metabolism are degraded by a pathway involving RNase III (Rnt1) and Rrp6 (Lee et al., 2005). A pathway termed DRN (degradation of mRNAs in the nucleus) was described that degrades a subset of mRNAs depending on the action of the cap-binding complex and Rrp6 (Kuai et al., 2005). Expression or processing of two mRNAs (*NRD1* and *CTH2*) is regulated by the nuclear RNA surveillance machinery and has been extensively discussed in Chapter 3.

In addition, many other mRNAs were frequently recovered with Trf4 (18% of all hits mapped to mRNAs), Nrd1 (21%) and Nab3 (31%) (Figure 3.4 A-D) suggesting either the presence of abundant cryptic sense RNAs or a very active regulated nuclear pre-mRNA turnover. Even distribution of the sense hits across many ORFs could be observed (representative examples in Figure 4.7 A and C). Therefore they most likely represent genuine mRNA targets and not the short promoter-associated transcripts that accumulate in strains lacking TRAMP and exosome activities (Davis and Ares, 2006; Wyers et al., 2005).

PGK1 and *TDH1* represent two highly expressed and frequently recovered mRNAs (Figure 4.7 A and C). High-throughput sequencing hits were distributed throughout the coding sequence of both mRNAs, rather than at the promoter, and all three factors associated in similar locations with the messages. No hits were mapped to the region downstream of the genes. Similar distributions of sequencing reads were seen for the majority of the mRNAs recovered (data not shown). Many of the recovered sequences also carried non-templated oligo(A) tails which were frequently in close proximity to Nrd1 and Nab3 consensus binding motifs (Figure 4.7 B and D). Oligoadenylated sequences were mapped to locations across the message, rather than around the cleavage and polyadenylation sites downstream of the ORFs, indicative of crosslinking to degradation intermediates. It is possible that previously undetected

but abundant, cryptic sense RNAs are responsible for these hits. However, I predict that this finding indicates that nuclear turnover of mRNA precursors is substantially more active than currently believed – and that the CRAC approach generates a snapshot of mRNA that are normally turned over rapidly in the cell nucleus.

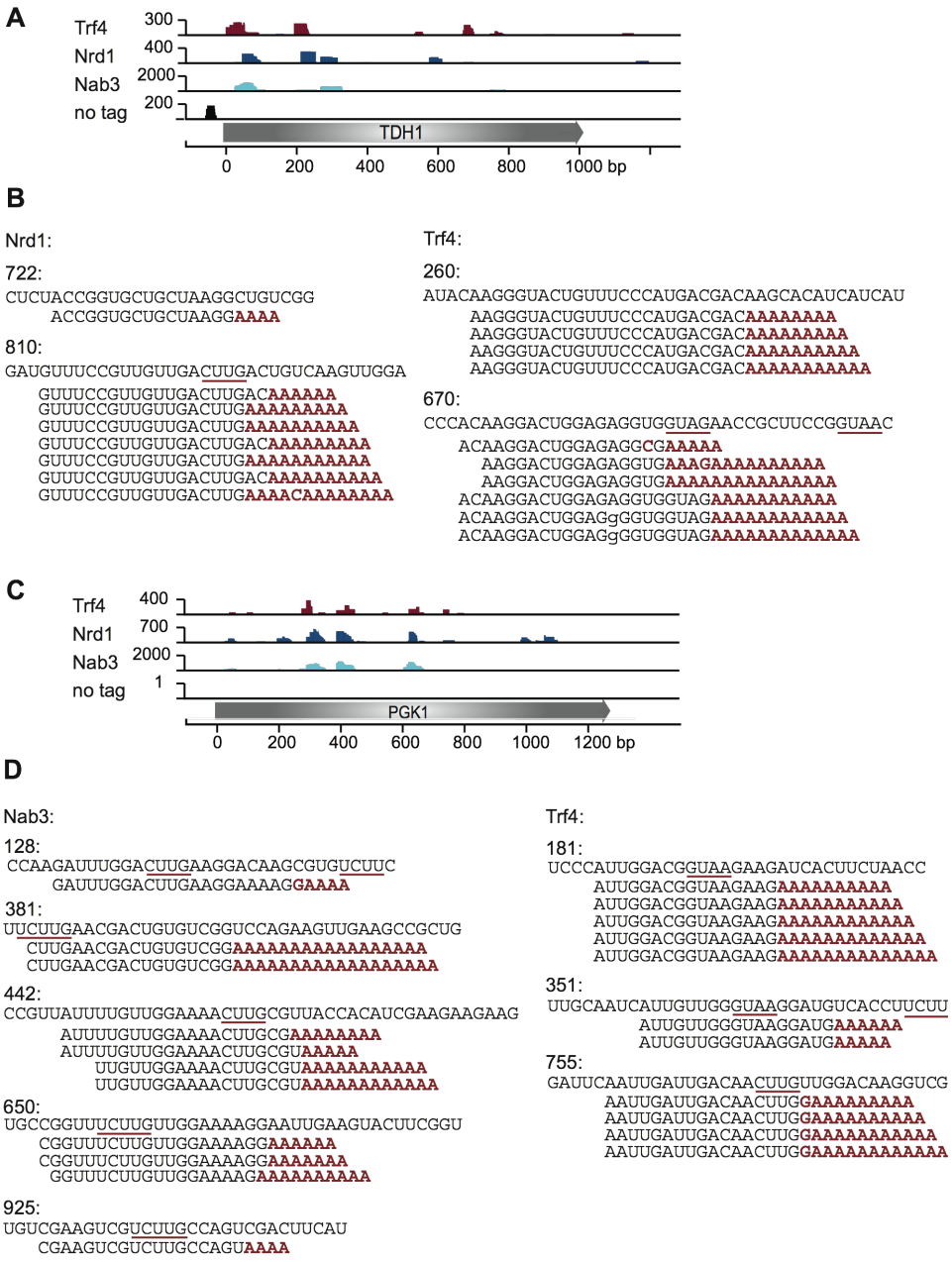

Figure 4.7 Crosslinking sites over PGK1 and TDH1
(A) and (C) Distribution of high-throughput sequencing reads of RNAs associated with the indicated proteins plotted over the *TDH1* (A) and *PGK1* (C) genes.
(B) and (D) Alignments of representative adenylated RNA reads from *TDH1* (B) and *PGK1* (D) mRNAs associated with the indicated proteins. Nucleotide positions are

given with respect to the 5' end of the gene. Deletions and mutations within the sequencing reads are shown in red. Nrd1 and Nab3 consensus binding motifs are underlined.

4.6 Discussion

These crosslinking results provide a genome-wide overview of the RNA population that is targeted by the nuclear RNA surveillance system in wild-type cells. Numerous known substrates of the Nrd1-Nab3 and TRAMP complexes were recovered including cryptic ncRNAs and antisense transcripts (Figure 4.1). The identification of the ncRNAs confirms that these are actively transcribed in wild-type cells, and not solely produced in response to deficient surveillance activities. These ncRNAs included the *GAL10as* RNA (Figure 4.2), which is present at around one molecule per 13 cells, supporting the sensitivity and reliability of the technique.

Identification of these transcripts is notable, because previous genome wide studies to identify cryptic or regulatory ncRNAs have only been carried out in the absence of components of the surveillance machinery (Neil et al., 2009; Wyers et al., 2005; Xu et al., 2009). In these studies tiling microarrays were used to identify novel ncRNAs that are stabilised under different nutrient conditions or in the absence of the nuclear exosome factor Rrp6. The tiling array and CRAC datasets each contained transcripts that were not identified with the other approach. Predicted CUTs that were not recovered in the CRAC analyses may be targeted for degradation by other nuclear surveillance factors, such as Rrp47 or Mpp6 (reviewed in Houseley and Tollervey, 2009). Conversely, many RNAs identified by sequencing during CRAC were overlooked in tilling array datasets. This will partly reflect differences in sensitivity between the data provided by microarrays compared to sequence-based data in measuring relative RNA expression. However, Figure 3.10 shows that most of the oligo(A) tails present on RNA surveillance substrates are too short to be recovered by oligo(dT) selection or to act as primers for oligo(dT) primed cDNA synthesis. Since these are important steps in previous microarray analyses, such RNAs could have been lost during sample preparation.

With the CRAC approach I was able to identify numerous, previously unknown transcripts that run antisense to mRNA genes. A few examples of asRNAs have been described that affect the expression of their sense mRNAs, either by transcription interference (Hongay et al., 2006) or by RNA-mediated transcriptional gene silencing (Berretta et al., 2008; Camblong et al., 2007; Houseley et al., 2008). In the latter cases, expression of the asRNAs aids/leads to recruitment of chromatin modifying complexes, which promote histone deacetylation and therefore silencing of the mRNA expression. Thus, asRNAs in yeast serve as an instrument to regulate gene expression despite the lack of the RNAi machinery. The identification of novel asRNAs is therefore very desirable in order to better understand the regulation of gene expression in yeast. The CRAC data for the surveillance factors provides this information, as the studied proteins interact with the asRNA population of a WT cell while it is being degraded. Hence, it can be used to identify new examples of regulated gene expression independent of artificial stress situations or a mutant background. Analysing the crosslinking of surveillance factors under different nutrient or stress conditions would potentially reveal how the asRNA population in the cell is altered in response to the changed environmental conditions.

The asRNAs *CAF17as* and *HPF1as* were stabilised *in vivo* in the absence of the nuclear surveillance machinery (*rrp6*Δ and *trf4*Δ, Figures 4.5 and 4.6). Moreover depletion of Nrd1 or Nab3 led to accumulation of read-through transcripts, which is in agreement with their reported role in transcription termination on snoRNAs and CUTs (Arigo et al., 2006b; Vasiljeva and Buratowski, 2006). However, it was previously believed that Nrd1 and Nab3 act only in transcription termination on short transcripts, as they associate with the CTD of Pol II when it is phosphorylated at Ser5, a mark characteristic of initiating polymerase and the 5' ends of genes (Vasiljeva et al., 2008a). The crosslinking data revealed association of Nrd1, Nab3 and Trf4 with the long *HPF1as* RNA and sequencing hits were evenly distributed over the first 4000 nt of the transcript. Furthermore, the stabilised read-through product was approximately 8000 nt long. These observations show that Nrd1 and Nab3 can associate with long RNAs and act in their transcription termination. This

argues against an exclusive function of Nrd1-Nab3 in association with the Ser5 phosphorylated Pol II CTD.

Northern blot experiments did not show that expression of mRNAs and asRNAs is mutually exclusive (Figures 4.5 and 4.6), as it would be expected if mRNA repression is always mediated by a transcription interference mechanism. There was also no strict anti-correlation in expression of *PHO84* mRNA and asRNA (Camblong et al., 2007). This asRNA is stabilised in aging cells or by loss of Rrp6 and mRNA levels decrease by a mechanism involving histone deacteylation. However, loss of *PHO84* mRNA transcription did not lead to asRNA stabilisation. Moreover, *PHO84* mRNA levels could be restored in the absence of the histone deacetylase, that mediates the silencing, or while the asRNA is expressed in the absence of Rrp6. This suggests that the regulated mRNA expression of *CAF17* and *HPF1* could be carried out in a similar fashion, utilising epigenetic marks that are dependent on the asRNA or its transcription.

A surprisingly large number of mRNAs were found to be directly bound by the surveillance factors tested. Oligoadenylated sequencing reads were found at various locations over the mRNA, identifying them as authentic targets for the RNA surveillance machinery. mRNAs encoding glycolytic enzymes (including *TDH2* and *PGK1* Figure 4.7) were frequently targeted by Nrd1, Nab3 and Trf4, and recovered sequences carried non-templated oligo(A) tails. This large group of mRNAs could conceivably reflect defective mRNAs that are targeted for degradation, but more likely suggests the presence of a nuclear mRNA turnover pathway.

Budding yeast lacks the miRNA systems present in most other eukaryotes analyzed. The miRNAs are believed to reduce the expression of large numbers of genes generally with only modest effects, although some stronger, more specific alterations are observed. I speculate that Nrd1-Nab3 and other surveillance factors may similarly act to modulate the expression of many genes, in addition to their strong effects on specific targets and defective RNAs.

4.7 Future plans-

Could Nrd1, Nab3 and Trf4 participate in regulated nuclear mRNA turnover?

The extent of the association of Trf4, Nrd1 and Nab3 with mRNAs encoding glycolytic enzymes was surprising, because the cultures for the crosslinking experiments were grown in complete yeast media containing glucose. Under these experimental conditions the yeast cells predominately produce energy by glucose fermentation, and therefore need glycolytic enzymes. However, prior to crosslinking, the cells were centrifuged, washed and resuspended in ice-cold PBS. This procedure takes approximately 30 minutes, allowing enough time for the yeast cells to adapt to the changed environmental conditions. Fast adjustment to varying growth conditions is essential for survival of a yeast cell outside the controlled laboratory environment. Therefore I speculate that during the preparation of the sample for crosslinking the yeast cells altered their metabolism, reflecting the reduced nutrient availability. In the sudden absence of available glucose the cells may systematically degrade mRNAs encoding for glycolytic enzymes. This is mediated, at least in part, by a nuclear pathway involving Nrd1-Nab3, TRAMP and most likely the exosome. Thus, the crosslinking data provide evidence for the existence of a regulated mRNA turnover pathway in the nucleus. In the future, I would like to further characterise this mRNA turnover pathway by assessing surveillance factor binding and turnover rates for newly synthesised mRNAs encoding glycolytic enzymes under different nutrient conditions. This could be done by combining CRAC with 4-thiouracil pulse labelling (see Chapter 6).

Chapter 5

Nrd1 and Nab3 participate in RNA Pol III transcript surveillance together with the TRAMP complex

5.1 Introduction

RNA Pol III transcribes a large number of small stable RNAs including 5S rRNA, tRNAs, the RNA components of RNaseP and the signal recognition particle, U6 snRNA, the snoRNA snR52 and RNA170, which is of unknown function. Primary Pol III transcripts are subjected to RNA processing and many undergo post-transcriptional nucleotide modifications (reviewed by Phizicky and Hopper, 2010). All of these maturation steps are overseen by quality control mechanisms. Compared to Pol II transcripts, quality control and surveillance systems are poorly understood for Pol III transcripts. The TRAMP and exosome complexes were previously shown to participate in surveillance of RNA Pol III transcripts, degrading 3' truncated 5S rRNA (5S*) and hypo-methylated $tRNA_i^{Met}$ (Kadaba et al., 2004; Kadaba et al., 2006; Schneider et al., 2007; Vanacova et al., 2005). In contrast, other defective tRNAs were shown to be degraded by the nuclear 5' to 3' exonuclease Rat1 (Chernyakov et al., 2008). The roles of Nrd1-Nab3 were predicted to be restricted to Pol II transcripts, due to the interactions between Nrd1 and the CTD region of the large subunit of Pol II (Conrad et al., 2000; Vasiljeva et al., 2008a). Recent genome wide Pol II ChIP analyses have confirmed that Pol II is largely absent from Pol III genes, and this is also the case for Nrd1 (Kim et al., 2010; Mayer et al., 2010).

5.2 Nrd1, Nab3 and Trf4 crosslink to Pol III transcribed RNAs

Unexpectedly, crosslinking experiments clearly showed an association of Nrd1 and Nab3 with Pol III transcripts (Figure 3.4 and Table 5.1). Over 30% of reads for Nrd1 and 17% of Nab3 reads corresponded to Pol III transcribed RNAs, many of which were precursor RNAs. For Trf4, 50% of reads mapped to Pol I transcribed rRNAs, while 10% corresponded to the Pol III transcribed 5S rRNA. The previously identified target $tRNA_i^{Met}$ was frequently recovered with Trf4, as was the elongator $tRNA^{Met}$.

	no tag	Nrd1	Nab3	Trf4
Total Pol I	37,19	8,34	6,56	50,75
Total Pol II	47,81	59,32	73,03	33,35
5S	2,7	1,63	0,53	10,01
tRNA	8,02	29,12	15,5	4,54
RPR1	0,21	0,18	0,08	0,06
SCR1	0,02	0,20	0,44	0,29
U6	0,06	0,02	0,02	0,13
snR52	0,07	0,13	0,14	0,15
RNA170	0	1,62	0	0
Total pol III	**11,44**	**31,31**	**17,40**	**15,85**

Table 5.1 Nrd1, Nab3 and Trf4 crosslink to Pol III transcripts
Percentage of mapped high throughput sequencing reads in each CRAC IP for different RNA Pol III transcripts and total Pol I and Pol II RNAs is displayed. Values represent averages of two biological replicates.

5.3 Nrd1, Nab3 and Trf4 associate with 5S rRNA surveillance intermediates

Anderson and colleagues showed that in the absence of Trf4 or the nuclear exosome factor Rrp6, a 3' truncated form of 5S rRNA (5S*) is stabilized, which is approximately 20 nt shorter than mature 5S (120 nt) (Kadaba et al., 2006). The 5S* species was shown to be oligoadenylated, dependent on Trf4.

Nrd1, Nab3 and Trf4 were each frequently associated with 5S sequences that terminated between nucleotides 50 and 100. These often but not exclusively carried oligo(A) tails, indicating that they represent degradation intermediates (Figure 5.2 A and B). In addition, oligoadenylated sequences were found at, and downstream of, the mature 3' end of 5S (Figure 5.1 C), probably representing precursors to the truncated species. 5S rRNA contains several consensus Nrd1-binding motifs and sequencing data revealed nucleotide substitution in the second, and deletions in the fourth GUAA/G motif (deleted nucleotides underlined), indicating direct Nrd1 binding at these positions. Nab3 and Trf4 also bound this region of 5S, with Nab3 crosslinking to the fourth GUAG motif (deleted nucleotide underlined), and Trf4

crosslinking to the third motif and the nucleotides downstream of the fourth motif GUAGUGUAGUGGG (deleted nucleotides underlined Figures 5.1 B and C). These crosslinking sites were recovered in both Solexa datasets. Since Nrd1 and Nab3 are RNA binding proteins, I speculate that they directly recognize defective Pol III transcripts and then recruit downstream surveillance factors such as TRAMP and the exosome.

In *trf4*Δ and *rrp6*Δ mutants, 5S* is accumulated, however, northern analysis did not reveal clear stabilization of any distinct, truncated 5S species following metabolic depletion of Nrd1 or Nab3 (Figure 5.1 D). This may reflect the redundancy observed in many yeast RNA surveillance pathways; reviewed in (Houseley and Tollervey, 2009)

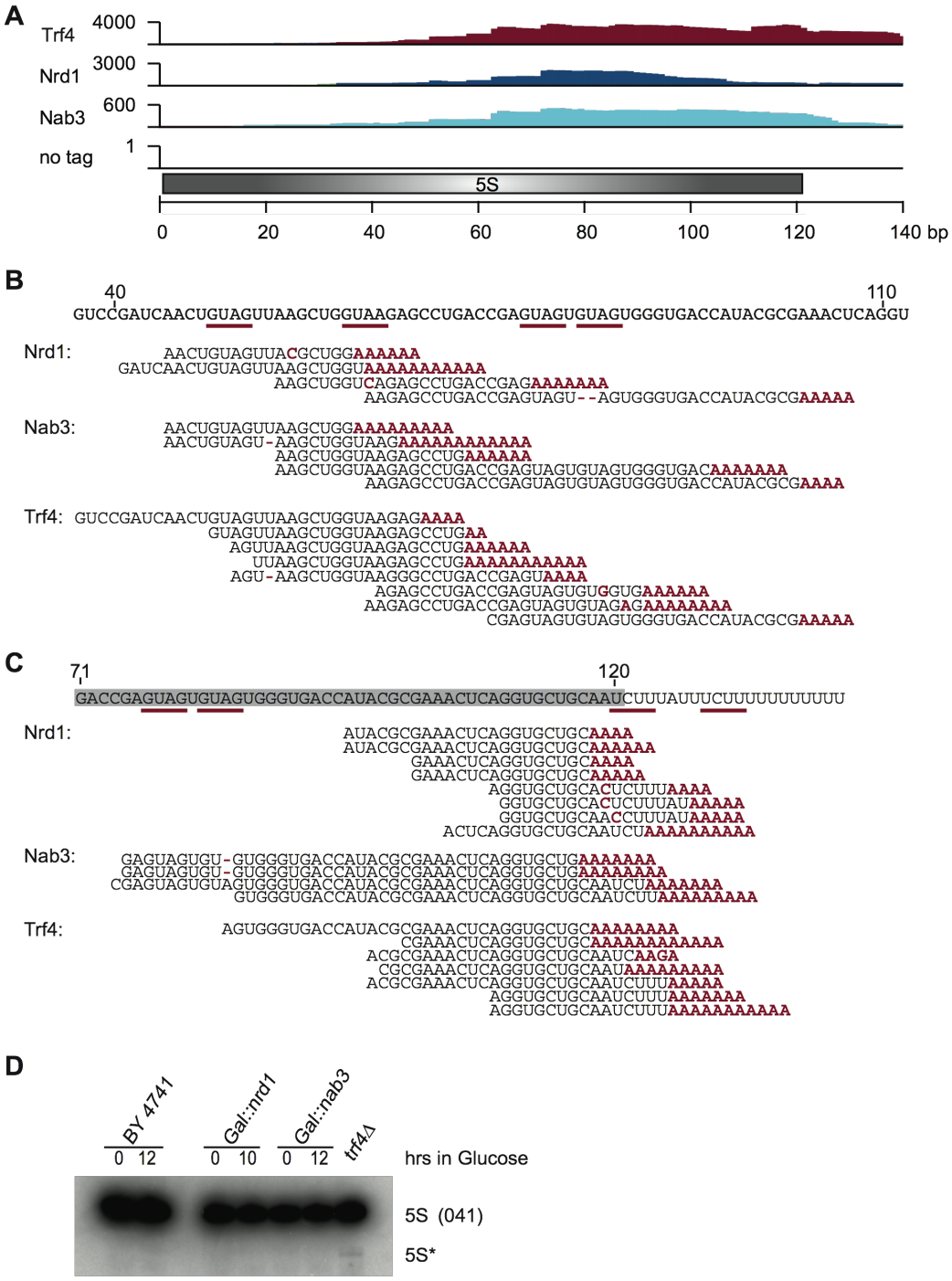

Figure 5.1 Truncated 5S rRNA is a target for the surveillance machinery
(A) High-throughput sequencing reads mapped to 5S pre-rRNA. Hits per one million reads are displayed.
(B) and (C) Alignment of representative high-throughput sequencing reads from the indicated IPs. Mature 5S is indicated by a grey box in (C). Numbers indicate nucleotide positions with respect to nucleotide +1 of 5S, consensus binding motifs are underlined, mismatches and deletions in sequencing reads are displayed in red.

(D) Northern blot of total RNA prepared from the indicated strains in exponential growth. P_{GAL}*::NRD1,* P_{GAL}*::NAB3* and BY4741 strains were pre-grown in YPGalSuc and then shifted to YPD for the indicated times. The *trf4*Δ strain was grown in YPD. RNAs were separated on 8% polyacrylamide/8M urea TBE gel and transferred to nylon membrane. Oligoprobes against the RNA species are indicated and the probe number is given in brackets.

5.4 Nrd1, Nab3 and Trf4 participate in surveillance of RNase P RNA precursor

RPR1 encodes the RNA component of RNase P and is transcribed by Pol III as a precursor containing a 5' leader and 3' trailer. Processing involves endonucleolytic cleavage and exonuclease digestion and requires RNP assembly; see (Srisawat et al., 2002) and references therein.

The CRAC experiments revealed the association of Nab3 and Trf4 with the 5' leader and Nrd1 and Trf4 with the 3' trailer of pre-*RPR1* (Figures 5.2 A and B). Sequence analysis for Nab3 identified deletions and therefore the exact crosslinking site within the 5' leader (Figure 5.2 B). Trf4 and Nab3 additionally bound sequences within the mature RPR1 (Figure 5.2 A and D). RNase P RNA forms a very compact structure (Figure 5.2 D), so the crosslinking sites for Nab3 and Trf4 at the 5' end, around position 100 and at 300 are in close proximity. It is conceivable that Nab3 and Trf4 associate with one region in the 3-dimensional structure. Deletions were also found in the Nab3 and Trf4 associated sequences, marked with * in the structure shown in Figure 5.2 D, identifying the crosslinking sites. Sequences recovered in the Nrd1 experiment did not contain the full 3' trailer up to the transcription stop site but carried extensions and non-templated oligo(A) tails (Fig. 5.2 B), suggesting the participation of Trf4 in marking this RNA for degradation. Oligo(A) tails were observed for Trf4 associated sequences mapping to the 3' extended form (Figure 5.2 B) and at internal sites (data not shown), which presumably represent surveillance intermediates. The same crosslinking and oligoadenylation sites were recovered in both Solexa datasets.

Therefore I predict that pre-*RPR1* can be recognized by Nrd1-Nab3 and oligoadenylated by Trf4 to mark it for exosome degradation, potentially as a consequence of sporadic defects in RNA folding and/or RNP assembly. To test this hypothesis, poly(A)$^+$ RNAs from *trf4*Δ as well as Nrd1 and Nab3 depleted cells were analyzed by northern blots. Pre-*RPR1* was detectably polyadenylated in WT cells (Figure 5.2 C), presumably reflecting normal surveillance activity. However, this polyadenylation was lost when Trf4 was absent or following depletion of Nrd1 or Nab3. The P_{GAL}*::NRD1,* P_{GAL}*::NAB3* strains showed reduced levels of pre-*RPR1* in the poly(A)$^+$ fraction, while the levels of mature *RPR1* and pre-*RPR1* levels remained constant in the total RNA of all tested strains. This demonstrates that the primary defect is not in *RPR1* processing but in surveillance of the, presumably defective, RNA. Hence, I conclude that Nrd1 and Nab3 are required to recognize defective pre-*RPR1* and to recruit Trf4 to mark it with an oligo(A) tail.

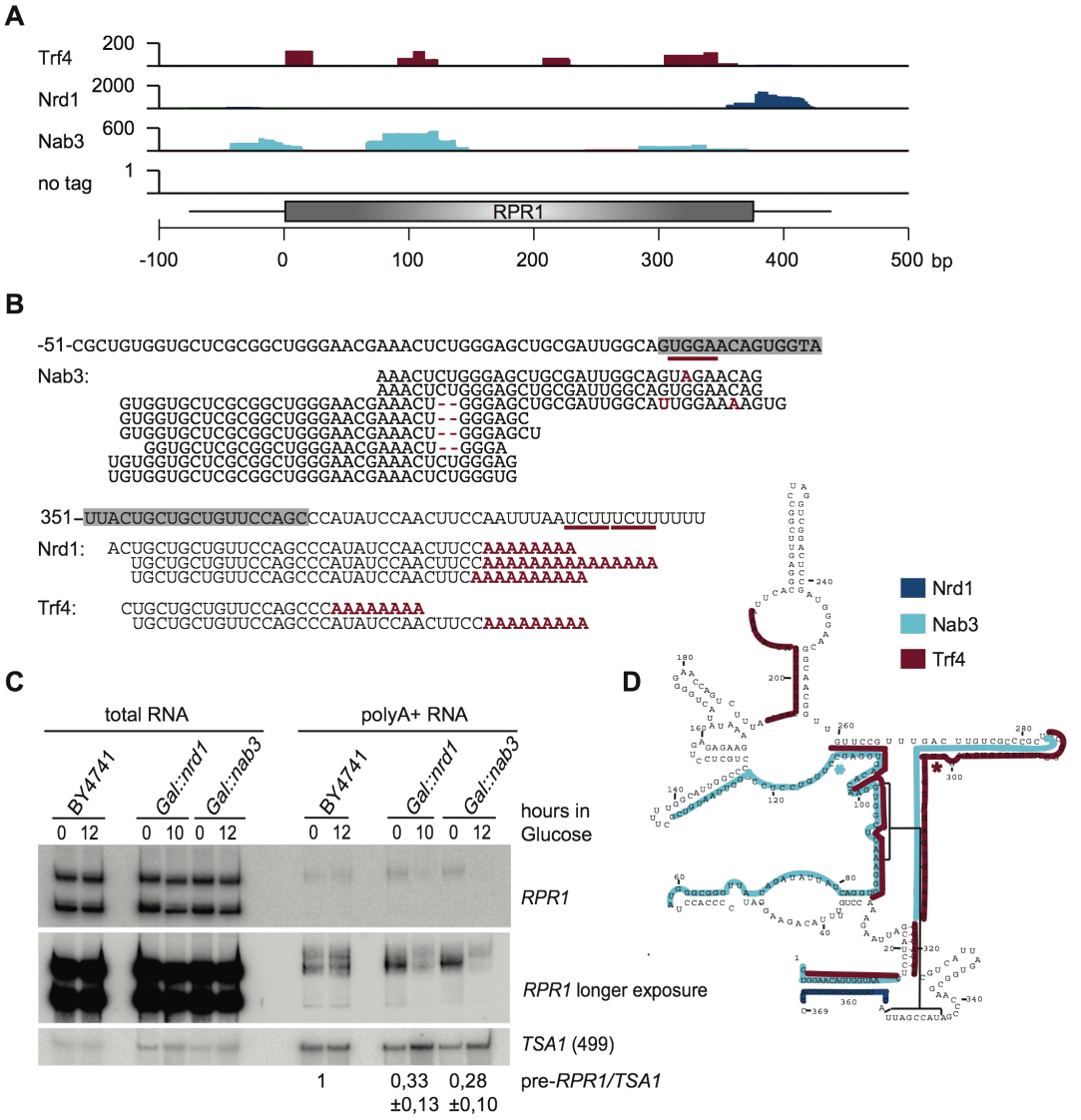

Figure 5.2 Pre-*RPR1* is polyadenylated and targeted by the surveillance machinery

(A) Densities of high-throughput sequencing reads mapped to pre-*RPR1*.

(B) Alignment of representative high-throughput sequencing reads from the indicated IPs to pre-*RPR1*. Grey boxes mark the mature *RPR1* sequence and numbering indicates the nucleotide position with respect to nucleotide +1 in the *RPR1* gene. Mismatches and deletions in sequencing reads are displayed in red.

(C) Northern analyses of total and poly(A)$^+$ RNA from P_{GAL}*::NRD1,* P_{GAL}*::NAB3* and BY4741 strains. Numbers below the northern report quantification of polyadenylated pre-*RPR1* relative to *TSA1* mRNA. The ratio of expression levels after 12 h in glucose compared to 0 h was set to 1 for the WT. Average of three biological replicates with standard deviations.

(D) Two-dimensional structure of the yeast RNase P RNA (Srisawat et al., 2002). High-throughput sequencing reads mapped to pre-*RPR1* are highlighted. Deletions in the sequencing reads are marked with *.

5.5 Nrd1 and Nab3 participate in pre-tRNA surveillance

Pre-tRNAs are transcribed with a 5' leader and 3' trailer, and in some cases they also contain introns. To generate a functional tRNA these must be removed and many base modifications are introduced (reviewed by Hopper and Phizicky, 2003; Phizicky and Hopper, 2010).

The CRAC data contained many examples of Nrd1, Nab3 and Trf4 associated pre-tRNAs. These generally contained introns (Figure 5.5 A), the 5' leader sequence (Figure 5.6 A) or 3' extensions (Figure. 5.3 A). Almost none of the recovered tRNAs carried the 3' CCA sequence that gets added as a late step in tRNA maturation. In contrast, many retained the 3' oligo(U) tract of the primary transcript followed by a non-encoded oligo(A) tail, or carried an oligo(A) tail following the coding region (Figure 5.3 B). The recovery of many oligoadenylated RNA that extend to the Pol III terminator indicated that tRNA-like species detected in CRAC analyses are derived from *bona fide* pre-tRNAs. The (pre-)tRNA regions that were recovered with Nrd1-Nab3 commonly carried consensus binding motifs, and mutations were frequently found within and around the GUAA/G and UCUU/CUUG sequences (underlined in Figure 5.3 B, 5.4 B and 5.6 B), indicating direct protein binding to these nucleotides.

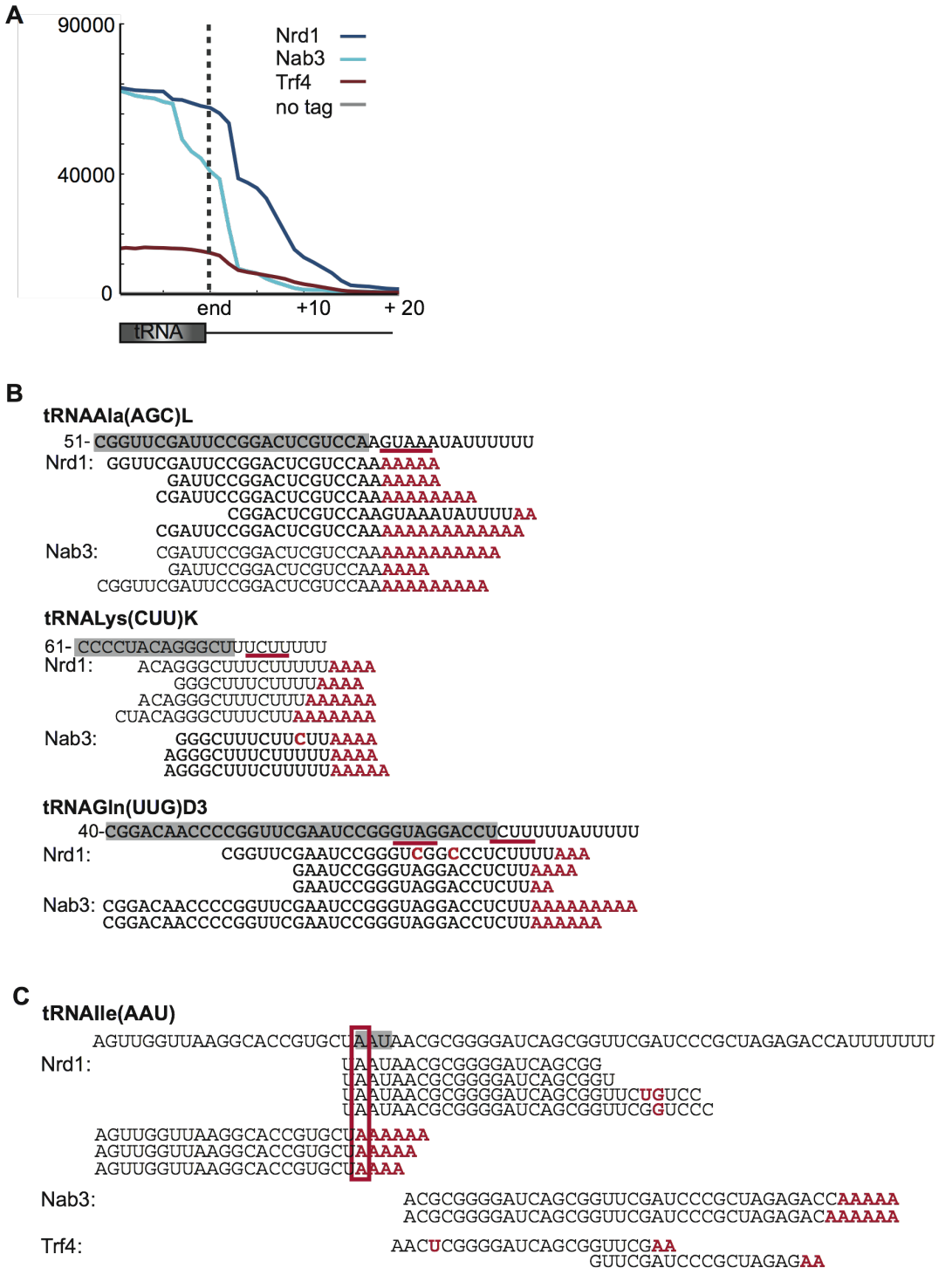

Figure 5.3 Nrd1 and Nab3 crosslink to 3' extended and oligoadenylated pre-tRNAs

(A) High-throughput sequencing reads of all tRNAs in the indicated IP are plotted with respect to the end of the tRNA.

(B) Alignment of representative high-throughput sequencing reads in the indicated IP to tRNAs. Grey boxes indicate the mature tRNA sequence and numbering gives the nucleotide position with respect to the first nucleotide of the tRNA gene. Mismatches

and deletions in the sequencing reads are displayed in red. Nrd1 and Nab3 consensus binding motifs are underlined.
(C) Alignment of representative high-throughput sequencing reads in the indicated IP to $tRNA^{Ile(AAU)}$. The grey box indicates the tRNA anticodon and adenosine 34 is marked with a red frame. Mismatches and deletions in the sequencing reads are displayed in red.

Further evidence for recovery of *bona fide* pre-tRNA sequences was the retrieval of pre-$tRNA^{Ile(AAU)}$ in the Nrd1 and Nab3 datasets. Adenosine 34 in the anticodon loop of $tRNA^{Ile(AAU)}$ is post transcriptionally edited to inosine (appearing as a G in the sequencing reaction), allowing the recognition of not only AUU, but also AUC and AUA mRNA codons (Auxilien et al., 1996). RNAs recovered from the CRAC experiments consistently contained the non-edited anticodon loop sequence (Figure 5.3 C), indicating that they originated from pre-tRNAs rather than mature tRNA. Further sequences mapping to $tRNAIle^{(AAU)}$ contained non-templated A tails, presumably representing surveillance intermediates.

I therefore speculate that (presumably defective) pre-tRNAs are bound by Nrd1-Nab3 and targeted for TRAMP-exosome degradation, similar to the pathway for Pol II transcripts.

An obvious possibility was that the tRNA-like RNAs observed in the Nrd1 and Nab3 CRAC datasets arose not from Pol III transcription, but from spurious Pol II transcription through the region. As described above (Chapters 3 and 4) such cryptic RNAs are expected to be targeted for Nrd-Nab/TRAMP/exosome dependent degradation. However, this appears unlikely as many of the recovered transcripts stopped at the Pol III terminator (Figure 5.3 B), indicative of authentic tRNA molecules.

Further evidence for Pol III transcription was provided by recovery of $tRNA^{Arg(ACG)}$. This tRNA is encoded by 6 genes; five of these have identical tRNA sequences but one carries two single nucleotide substitutions ($A_{43}G$ and $A_{58}G$; encoded by the gene *tR(ACG)J*; Figure 5.4 A). The CRAC analysis of Nrd1, Nab3 and Trf4 preferentially recovered the $tRNA^{Arg(ACG)J}$ pre-tRNA, frequently with oligo(A) tails (Figure 5.4 D).

RNA folding algorithms predict that the $tRNA^{Arg(ACG)J}$ variant is less likely than the major form of $tRNA^{Arg(ACG)}$ to fold into the correct $tRNA^{Arg}$ structure (Figure 5.4 B and C). The major $tRNA^{Arg(ACG)}$ can adopt two conformations of similar predicted free energy, a stem loop or the clover leaf. Two single nucleotide substitutions in $tRNA^{Arg(ACG)J}$ are predicted to greatly favor the stem loop structure, providing a clear rationale for its targeting by the surveillance system. These differences in folding would not, however, have been predicted to alter the fate of a spurious RNA Pol II transcript.

Primer extension analysis with oligos discriminating the two different $tRNA^{Arg(ACG)}$ species (Figure 5.4 E) detected the primary transcripts for both tRNAs. However, for $tRNA^{Arg(ACG)J}$ no mature transcript was detected, whereas for the major arginine tRNA the mature species, with stops corresponding to all post-transcriptional modifications, was readily identified. I conclude that pre-$tRNA^{Arg(ACG)J}$ is synthesized by Pol III, but fails to fold correctly and is efficiently degraded by the surveillance machinery. Depletion of either Nrd1 or Nab3 did not restore levels of mature $tRNA^{Arg(ACG)J}$ (Figure 5.4 E). This suggests that the pre-tRNA is degraded by a redundant pathway, most likely from the 5' end by Rat1 as shown for other defective tRNAs (Chernyakov et al., 2008). Preliminary primer extension experiments carried out with RNA prepared from yeast strains carrying mutations in the 5' to 3' RNA degradation pathway (*xrn1*Δ*, rai1*Δ and *Met::rat1*; RNAs were a generous gift from Dr. Aziz ElHage) showed slight stabilisation of the mature $tRNA^{Arg(ACG)J}$ (data not shown). Redundancy in $tRNA^{Arg(ACG)J}$ degradation could potentially be uncovered by double mutants in the 5' and 3' degradation pathway (P_{GAL}*::NRD1 rai1*Δ), However, attempts to generate this strain have proved unsuccessful.

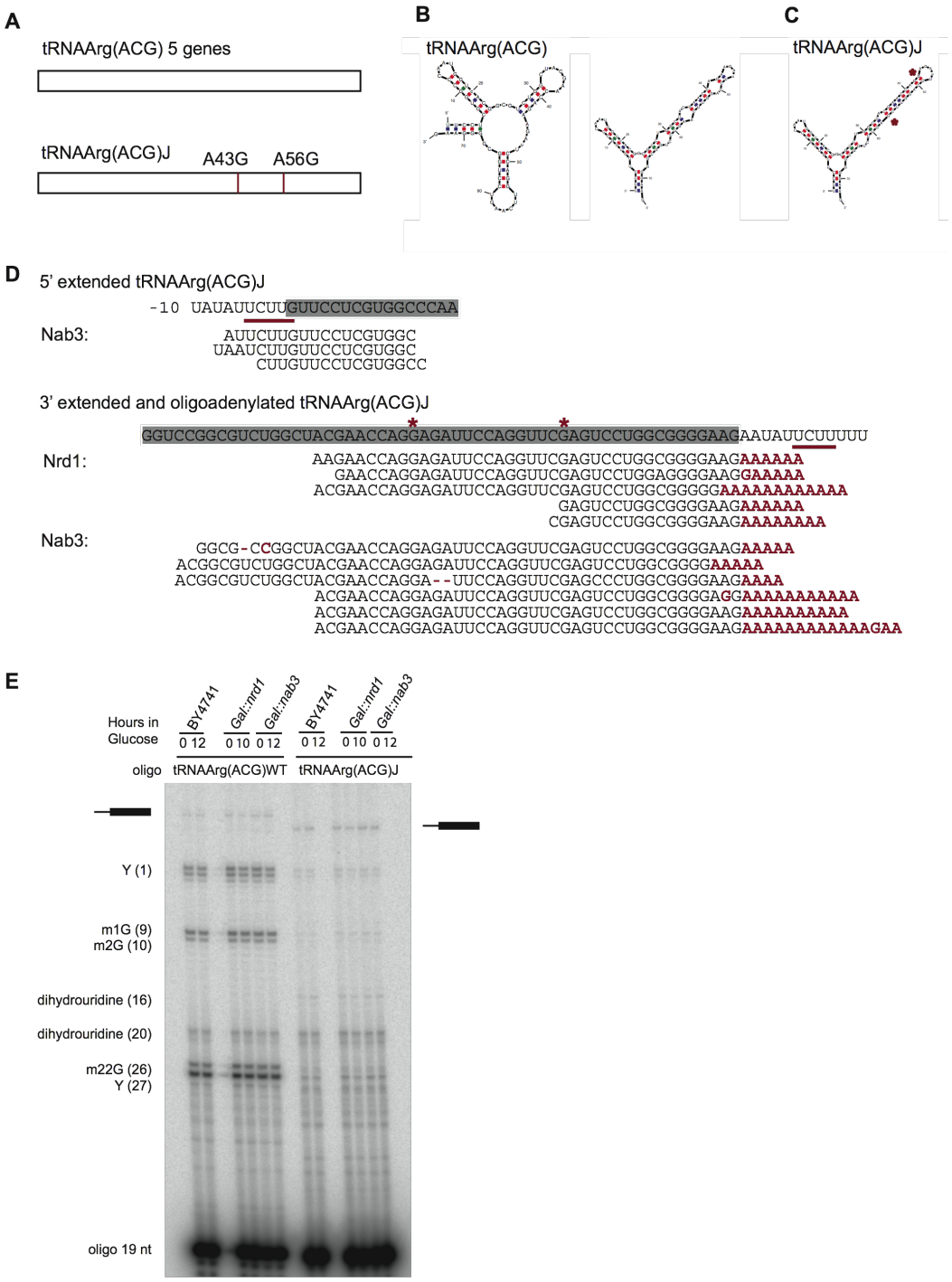

Figure 5.4 Surveillance of tRNA$^{Arg(ACG)J}$

(A) Model representing two endogenous tRNA^{Arg} species. Exchanged nucleotides are indicated.

(B) tRNA folding predicted by the Zucker algorithm for the major form of tRNA$^{Arg(ACG)}$ with two energetically equivalent conformations.

(C) Fold predicted for the tRNA$^{Arg(ACG)J}$ variant . Exchanged nucleotides are indicated with *.

(D) Alignment of representative high-throughput sequencing reads associated with the indicated proteins (on the left) to the pre-tRNA$^{Arg(ACG)J}$. Grey boxes indicate the mature tRNA sequence and numbering indicate the nucleotide position with respect to the first nucleotide of the tRNA gene. Mismatches and deletions in the sequencing reads are displayed in red. Exchanged nucleotides are indicated with *. Nrd1 and Nab3 consensus binding motifs are underlined.
(E) Primer extension of total RNA prepared from the indicated strains in exponential growth. *P_{GAL}::NRD1, P_{GAL}::NAB3* and BY4741 strains were pre-grown in YPGalSuc and then shifted to YPD for the indicated times. Primer extension products were separated on 12% polyacrylamide/8M urea TBE gels. Oligos specifically hybridizing to the major or variant forms of tRNA$^{Arg(ACG)}$ were used as indicated above. RNA modifications causing a primer extension stop are indicated on the left, nucleotide position from the tRNA start is given in brackets and position of the primary tRNA transcript is indicated on the top of the gel.

To further assess the participation of Nrd1-Nab3 in pre-tRNA surveillance, strains depleted for Nrd1 or Nab3 were analyzed by northern hybridization (Figure 5.5 B). Five different intron-containing tRNAs that were identified in CRAC analyses as targets for Nrd1 and Nab3 binding were tested. Three biological replicates were performed and the pre-tRNA:tRNA ratio was determined (Figure 5.5 D). Each showed reproducible accumulation of the unspliced pre-tRNA in strains depleted of Nrd1 or Nab3. The increase of pre-tRNA species was not accompanied by loss of the mature tRNA, indicating that the pre-tRNA accumulation does not reflect a processing defect. Steinmetz and colleagues (Steinmetz et al., 2001) noted that genes encoding components of the tRNA splicing machinery lie downstream of snoRNA genes. This suggested that their expression might be reduced in *nrd1* and *nab3* mutants due to transcription read-through, leading to pre-tRNA splicing defects. To test this hypothesis, I compared the phenotypes of the tRNA-splicing endonuclease mutant (*P_{GAL}::SEN34*) with the metabolic depletion of Nrd1 or Nab3 (Figure 5.5 C). Depletion of the tRNA splicing endouclease Sen34 lead to stronger accumulation of pre-tRNAs and loss of mature tRNA, as might be expected for an RNA processing mutant. These differences support the model that Nrd1 and Nab3 act in surveillance.

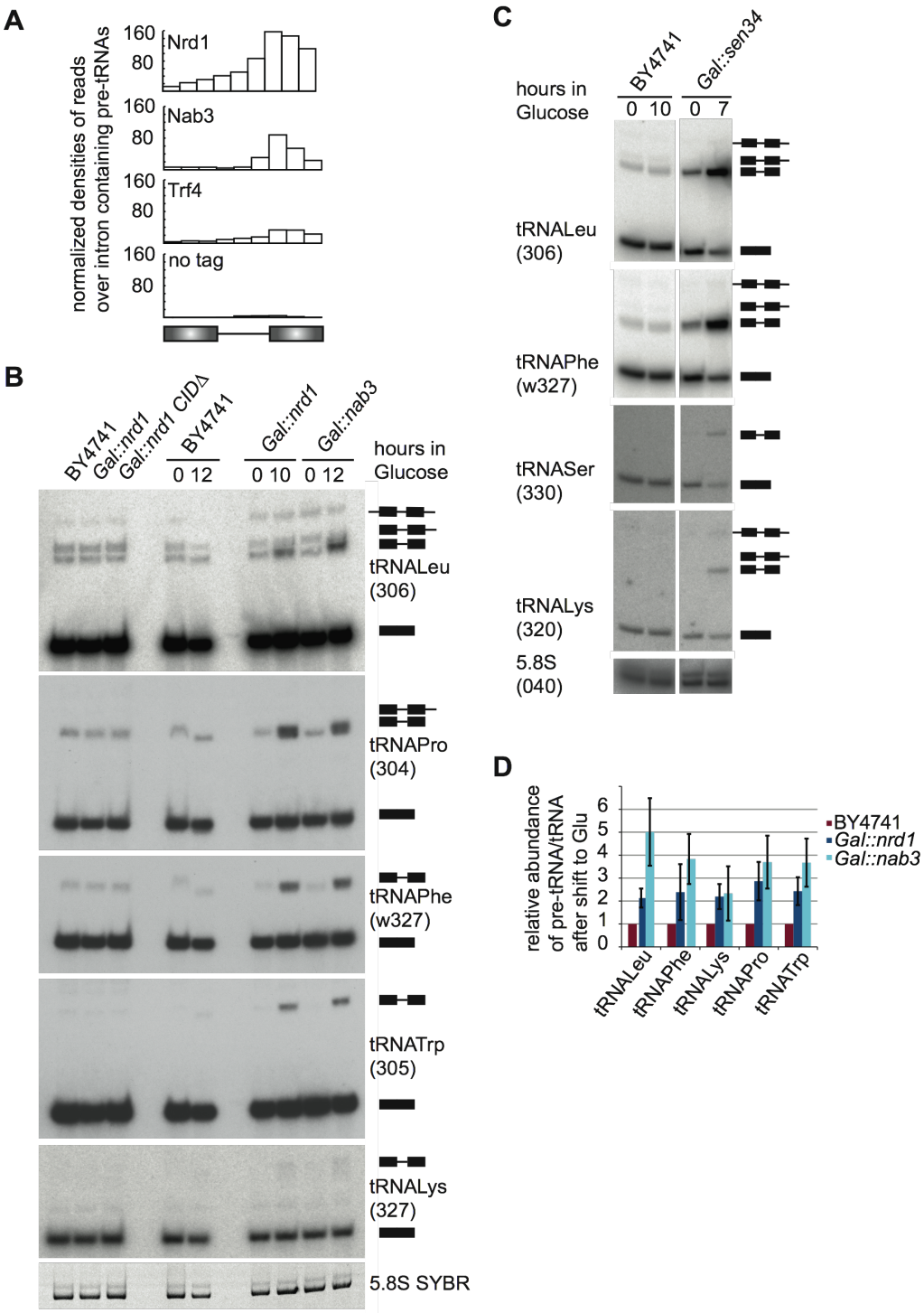

Figure 5.5 Nrd1 and Nab3 participate in surveillance of intron containing pre-tRNAs

(A) Average densities of reads mapped to intron-containing tRNAs. tRNA exons and introns have various lengths; all exons and introns were divided into three bins and density of reads in each bin is displayed.

(B) and (C) Northern blot of total RNA prepared from the indicated strains in exponential growth. *P_{GAL}::NRD1, P_{GAL}::NAB3, P_{GAL}::SEN34* and BY4741 strains were pre-grown in YPGalSuc and then shifted to YPD for the indicated times.

BY4741, P_{GAL}*::NRD1 and* P_{GAL}*::nrd1CID*Δ strains were grown in YPGalSuc prior to RNA extraction. RNAs were separated on 8% polyacrylamide/8M urea TBE gel and transferred to nylon membrane. Oligo probes directed against different tRNA species are indicated and number of the probe is given in brackets. A schematic representation of the identified species is displayed. 5.8S rRNA serves as a loading control.
(D) Quantification of pre-tRNA relative to mature tRNA expression after metabolic depletion of Nrd1 and Nab3 compared to time point 0 in (B). Expression levels were set to 1 in the WT and given as an average of three biological replicates with standard deviations.

To rule out the possibility that the tRNAs identified by CRAC represent spurious Pol II transcripts, and to demonstrate that the function of Nrd1 and Nab3 in Pol III transcript surveillance is Pol II independent, I analyzed a *nrd1* mutant that is unable to interact with the CTD of Pol II (*nrd1CID*Δ; Vasiljeva et al., 2008a). In this mutant, levels of mature and pre-tRNAs were unchanged (Figure 5.5 B), demonstrating that the observed effects of Nrd1 and Nab3 on tRNAs are indeed independent of their association with RNA polymerase II.

Precursor tRNAs carrying 5' extensions were also identified among the CRAC targets for Nrd1 and Nab3 (Figure 5.6 A and B). Northern hybridization validated these crosslinking results as increased levels of 5' extended $tRNA^{Arg(UCU)}$ were seen following depletion of Nrd1, and even more so of Nab3, whereas levels of mature $tRNA^{Arg(UCU)}$ were not altered. Quantification of RNA levels is shown for three biological replicates. Consistent with the observation made for the intron containing pre-tRNAs, the accumulation of 5' extended pre-tRNAs is most likely not a result of a processing defect, since the 5' trailer is cleaved by RNase P and mature *RPR1* RNA was formed in P_{GAL}*::NRD1* and P_{GAL}*::NAB3* strains (compare Figure 5.2 C).

Thus, I conclude that pre-tRNAs with defects in folding or maturation are bound by Nrd1-Nab3 and targeted to the TRAMP-exosome degradation pathway.

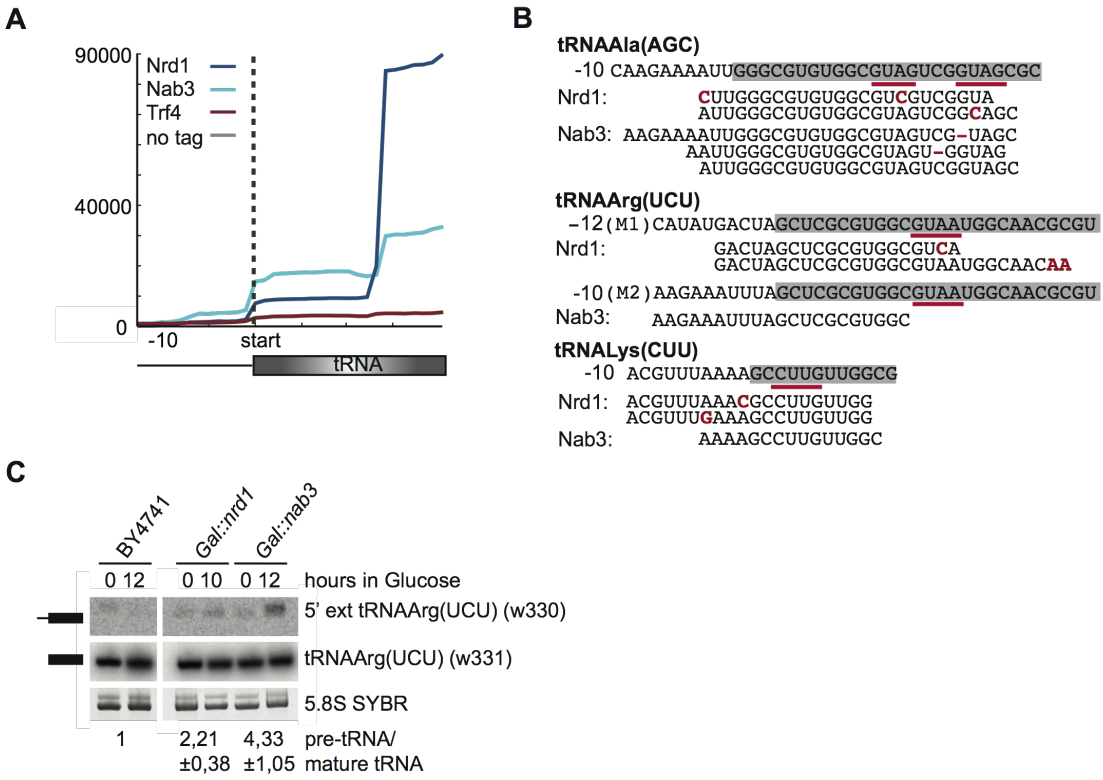

Figure 5.6 Nrd1 and Nab3 participate in surveillance of 5' extended tRNAs
(A) High-throughput sequencing reads of all tRNAs associated with the indicated proteins are plotted with respect to the start of the tRNA. Position 0 indicates the first nucleotide in the mature tRNA.
(B) Alignment of high-throughput sequencing reads of RNAs associated with the indicated proteins (on the left) on 5' extended tRNAs. Grey boxes mark the mature tRNA sequence and numbering specifies the nucleotide position with respect to the first nucleotide of the mature tRNA. Mismatches and deletions in the sequencing reads are displayed in red. Nrd1 and Nab3 consensus binding motifs are underlined.
(C) Northern blot of total RNA prepared from the displayed strains in exponential growth. *P_{GAL}::NRD1, P_{GAL}::NAB3* and BY4741 strains were pre-grown in YPGalSuc and then shifted to YPD for the indicated times. RNAs were separated on 10% polyacrylamide/8M urea TBE gel and transferred to nylon membrane. Oligo probes used are indicated and the number of the probe is given in brackets. 5.8S rRNA serves as a loading control. Quantification of the expression levels of the 5' extended tRNA relative to the mature tRNA is displayed below. The ratio of expression after 12 h compared to the 0 h time point it set to 1 for the WT and given as an average of three biological replicates with standard deviations.

5.6 Nrd1 and Nab3 may act together with TRAMP and exosome in pre-U6 surveillance

Spliceosomal U6 snRNA is transcribed by RNA Pol III and was associated with Nrd1, Trf4 and, to a small extent, with Nab3 in the CRAC experiment (Table 5.1 and Figure 5.7 A). U6 sequences recovered with Trf4 largely contained non-templated oligo(A) tails in both experiments. These were found a few nucleotides upstream of the mature 3' end, at the 3' end and on the 3' extended primary U6 transcript (Figure 5.7 B). Nab3 associated sequences were also 3' extended and contained the Pol III specific oligo(U) tract, but lacked non-templated adenosine residues.

Libri and colleagues reported that oligoadenylated U6 accumulates in the absence of the nuclear exosome component Rrp6, dependent on Trf4 polyadenylation activity (Wyers et al., 2005). To test whether Nrd1 or Nab3 influence U6 oligoadenylation, poly(A)$^+$ RNAs from *trf4*Δ and cells depleted for Nrd1 or Nab3 were analyzed by northern hybridization (Figure 5.7 C). U6 RNA was not detectably polyadenylated in WT cells at time point 0 h or in *trf4*Δ strains. In contrast, polyadenylation was observed in cells overexpessing Nrd1 or Nab3 (0 h in glucose), which was lost after depletion of the proteins. This suggests a similar role for Nrd1 and Nab3 in RNA recognition of, presumably faulty, (pre-)U6 and recruitment of TRAMP. This resembles the conclusions drawn above for pre-*RPR1*. However, given the low amount of polyadenylated U6 in WT cells, conclusions on the requirements of Nrd1 and Nab3 in pre-U6 recognition must be drawn with caution. Reduced levels of mature U6 snRNA were observed in *nrd1* and *nab3* mutants (Figure 5.7 C; total RNA, left lanes), but this phenotype was not very consistent between experiments (four biological replicates). Thus, few clear conclusions concerning the influence of Nrd1 and Nab3 on U6 expression or surveillance can be drawn at this point.

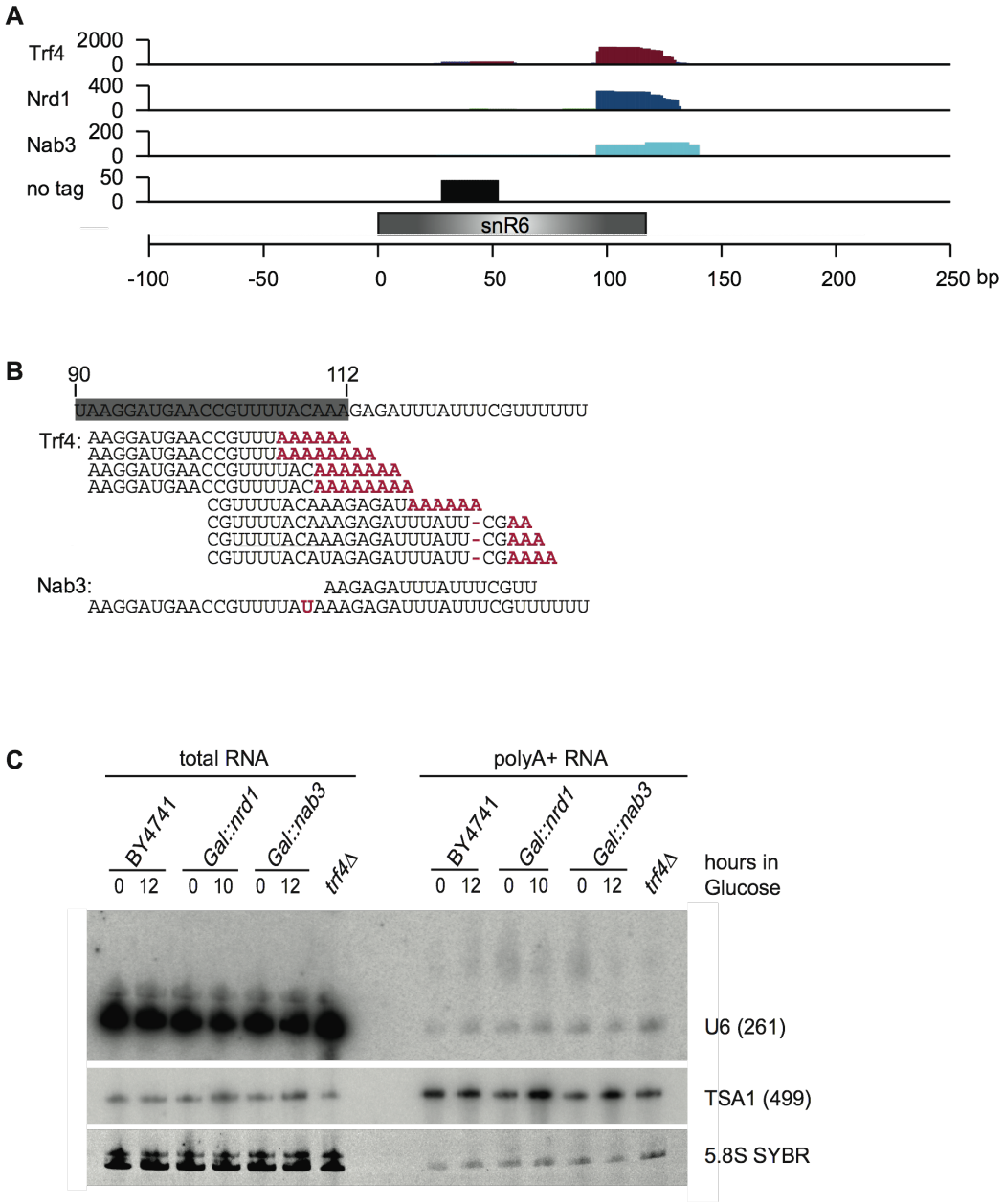

Figure 5.7 Nrd1 and Nab3 maybe participate in surveillance of pre-U6 together with Trf4

(A) High-throughput sequencing reads of all RNAs in the indicated IP are plotted over U6 snRNA.

(B) Alignment of representative high-throughput sequencing reads associated with the indicated proteins on (pre-)U6. Grey boxes mark the mature tRNA sequence and numbering specifies the nucleotide position with respect to the first nucleotide of the U6 gene. Mismatches and deletions in the sequencing reads are displayed in red. Nrd1 and Nab3 consensus binding motifs are underlined.

(C) Northern blot of total RNA (left) and poly(A)$^+$ RNA (right) prepared from the indicated strains in exponential growth. *P_{GAL}::NRD1, P_{GAL}::NAB3* and BY4741 strains were pre-grown in YPGalSuc and then shifted to YPD for the indicated times. *trf4*Δ was grown in YPD. RNAs were separated on 8% polyacrylamide/8M urea

TBE gel and transferred to nylon membrane. Oligo probes are indicated and number of the probe is given in brackets. *TSA1* mRNA and 5.8S rRNA serve as loading controls.

Previous studies on TRAMP and exosome mediated surveillance of Pol III transcripts demonstrated that a mutant form of U6, deleted for nucleotides 59-72, is stabilized in the absence of Rrp6 or Trf4 (Kadaba et al., 2006). To further characterise the function of Nrd1 and Nab3 in U6 surveillance, I investigated the role of Nrd1-Nab3 in surveillance of this U6 mutant. Yeast strains expressing plasmid borne *SNR6* (which encodes U6) under its own promoter (pBS1191 WT U6; pRS317 containing *SNR6* with 450 nt upstream and downstream; (Luukkonen and Seraphin, 1998) or *SNR6* Δ59-72 (pBS1191 U6 Δ59-72, this study Figure 5.8 A) were constructed. Expression of U6 was monitored by northern hybridisation after depletion of Nrd1, Nab3 or Trf4 (Figure 5.8 B and C). All strains also carried the endogenous copy of the *SNR6* gene, as the 59-72 deletion is otherwise lethal. Wild-type strains carrying the mutant allele grew like an otherwise isogenic strain bearing the WT U6 on a plasmid (Figure 5.8 B). Nrd1 and Nab3 depleted in the presence of the U6 mutant with the same growth kinetics as previously observed (compare Chapter 4). In contrast, expression of the U6 mutant strongly inhibited growth in a *trf4*Δ strain, with doubling times increased from ~3 h in *trf4*Δ to almost 10 h in the *trf4*Δ +pBS1191 U6Δ59-72.

Northern hybridisation showed expression of U6 snRNA in all strains, with stabilisation of the shorter form of U6 *trf4*Δ cells but not in the absence of Nrd1 or Nab3. Expression of full length U6 was only slightly increased in the BY4741 WT stain by the additional *SNR6* copy on the plasmid. Expression of U4 snRNA was not altered in any of the strains and serves as a control.

I conclude that Nrd1 and Nab3 might be involved in surveillance of pre-U6 snRNA but do not influence the stability of a specific, artificial U6 mutant.

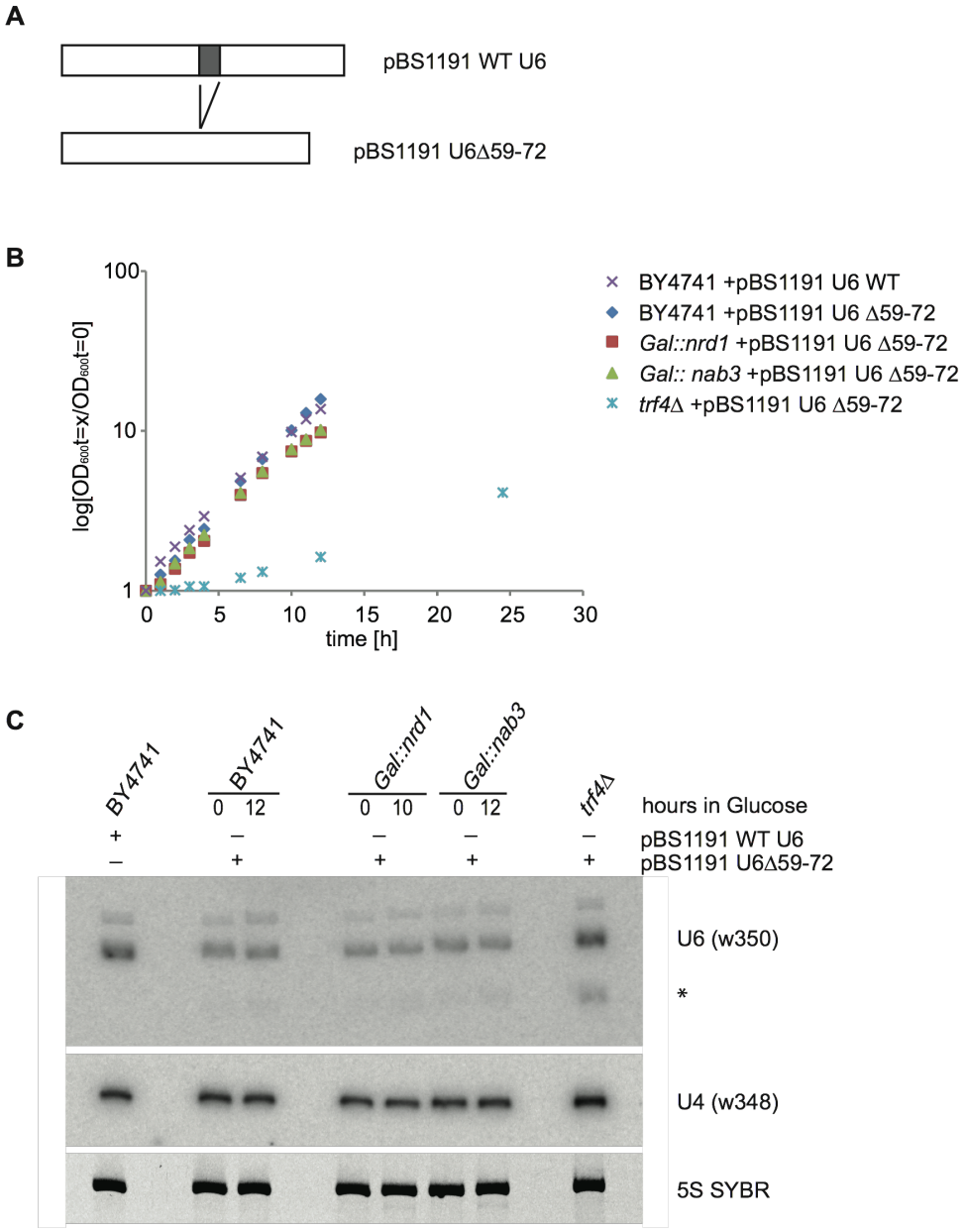

Figure 5.8 Nrd1 and Nab3 do not participate in surveillance of a specific U6 mutant

(A) Schematic representation of the U6 gene and the U6Δ59-72 mutant.

(B) Growth curve of the indicated strains expressing either plasmid borne pBS1191 U6 WT or mutant pBS1191 U6Δ59-72. *PGAL::NRD1, PGAL::NAB3* and BY4741 strains carrying pBS1191 U6Δ59-72 were pre-grown in SD GalSuc –URA and then shifted to SD Glu -URA for the indicated times. BY4741 carrying the WT plasmid and *trf4*Δ +pBS1191 U6Δ59-72 were grown in SD Glu -URA.

(C) Northern blot of total RNA prepared from the indicated strains in exponential growth. Strains were grown as in (B). RNAs were separated on a 10% polyacrylamide/8M urea TBE gel and transferred to nylon membrane. Oligo probes used are indicated and number of the probe is given in brackets. 5S rRNA serves as a loading control.

5.7 Discussion

The CRAC approach led to the surprising finding that the RNA Pol II associated factors Nrd1 and Nab3 were crosslinked to transcripts generated by RNA Pol III. In different analyses, between 15 and 30% of recovered sequences corresponded to Pol III RNAs (Table 5.1), with the majority of these being precursor RNAs or recognisably defective transcripts (Figures 5.1- 5.6). Oligoadenylation at internal sites or at the 3' end was also commonly observed in the sequence data (Figures 5.1 C, 5.2 B, 5.3 B, 5.4 D and 5.7 B), which is a clear indication of TRAMP-mediated surveillance. The association of Nrd1 and Nab3 with primary tRNA transcripts carrying a 3' oligo(U) tail (Figure 5.3 B), as well as with the variant $tRNA^{Arg(ACG)J}$ (Figure 5.4), indicates that these targets are authentic RNA Pol III transcripts and not cryptic Pol II transcripts. Recent genome wide Pol II ChIP analyses have shown that both Pol II and Nrd1 are largely absent from Pol III genes (Kim et al., 2010; Mayer et al., 2010). The TRAMP and exosome complexes were known to function in surveillance of some defective Pol III transcripts (Kadaba et al., 2004; Kadaba et al., 2006; Schneider et al., 2007; Vanacova et al., 2005) but the participation of Nrd1 and Nab3 in Pol III RNA surveillance was unexpected.

The association of Nrd1 and Nab3 *in vivo* with the known TRAMP substrate 5S* strongly indicated their involvement in the Pol III transcript surveillance pathway. Further support was provided by identification of the Nrd1 crosslinking site within 5S as the previously characterised *in vitro* binding motif GUAG (Figure 5.1 B). P_{GAL}*::NRD1* and P_{GAL}*::NAB3* mutants failed to clearly stabilise the truncated 5S species, in contrast to mutants of the TRAMP or exosome complexes. They are therefore able to associate with the aberrant 5S* *in vivo* but are dispensable for its degradation. It is likely that Nrd1 and Nab3 enhance recognition and degradation of 5S* by TRAMP and the exosome, but if Nrd1 and Nab3 are absent these factors may degrade the aberrant 5S* alone. Previous analyses showed that both the TRAMP complex and the exosome component Rrp44 could independently recognise a defective pre-$tRNA^{iMet}$ (Schneider, 2007). Only correctly matured 5S rRNA can function within the ribosome, so any aberrant form needs to be removed in order to

guarantee translation fidelity, and surveillance by parallel, redundant pathways may help to ensure this.

Crosslinking was observed to many Pol III RNA precursors, leading to the hypothesis that Nrd1-Nab3 function in general surveillance of Pol III transcripts. Oligoadenylated sequences were frequently recovered with both Nrd1 and Nab3, indicating that they act together with the TRAMP and exosome complexes and remain bound to target RNAs during oligoadenylation. Association with oligoadenylated precursor RNAs was observed for several tRNAs as well as for RNase P RNA *RPR1* and U6 snRNA (Figures 5.2 B, 5.3 B, 5.4 D and 5.6 B). I could demonstrate that loss of Nrd1 or Nab3 led to accumulation of unspliced or 5' extended form of pre-tRNA species that were identified as crosslinked to Nrd1 or Nab3 (Figures 5.5 B and 5.6 C). Pre-*RPR1* was shown to be oligoadenylated by Trf4, presumably as a consequence of misfolding or other structural defects, and this was dependent on recognition by Nrd1 and Nab3 (Figure 5.2 C).

The observation that Nrd1 and Nab3 were associated with the same sequences with and without poly(A) tails indicates that they act upstream of TRAMP in the recognition of target RNA. However, the presence of oligo(A) tails shows that they remain associated while Trf4 acts on the target RNA. It is plausible that Nrd1-Nab3 recognise and bind target RNAs, recruit Trf4 to mark them by oligo(A) addition and remain associated until the exosome executes degradation or processing. In this scenario the Nrd1-Nab3 complex could act as a marker, which constantly signals to different factors of the surveillance machinery that the bound RNA is a target.

Vasiljeva and colleagues made a similar observation when they studied Nrd1-Nab3 and exosome mediated degradation and processing *in vitro* (Vasiljeva and Buratowski, 2006). They found that the exosome is not able to fully degrade or process a 3' extended *GAL7* substrate containing Nrd1 consensus binding sites in the presence of the Nrd1-Nab3 complex. Hence, they suggested that the Nrd1-Nab3 complex acts as a 'road-block' forcing the exosome to stop. These observations support the model that Nrd1 and Nab3 stay attached to an RNA target until it is fully matured or degraded by the exosome complex.

Not all of the recovered RNA Pol III transcripts carry previously defined Nrd1 or Nab3 consensus binding motifs, unlike 5S or the displayed tRNAs (Figures 5.1, 5.3 and 5.6). Many tRNA targets do not carry a clear binding motif e.g. $tRNA^{Arg(ACG)}$, $tRNA^{iMet(CAU)}$ or $tRNA^{Phe(GAA)}$, while *RPR1* contains only one GUAG and one UCUU motif. Nevertheless, these transcripts are recognised by Nrd1 and Nab3 *in vivo* and pre-*RPR1* is subject to TRAMP/exosome mediated surveillance (Figure 5.2 C). This suggests that Nrd1 and Nab3 binding might not only be triggered through the characterised sequence motifs.

All Pol III transcripts fold into tight tertiary structures, potentially providing a different basis for recognition of defective structure. Folding into the characteristic L-shape of tRNAs is dependent on interactions between modified bases as well as the ribose-phosphate backbone. Lack of correct modification, removal of intronic sequences or preferential folding into an alternative structure, as shown for $tRNA^{Arg(ACG)J}$, may lead to deviation from the normal structure. I speculate that recognition may primarily reflect folding into an incorrect and more open structure, which could be monitored by Nrd1-Nab3 binding. This would give rise to a much more general recognition mechanism than sequence-specific binding.

Overall the CRAC results and validation of targets by northern hybridisation provide strong evidence that Nrd1-Nab3 participate together with the TRAMP and exosome complexes in surveillance of precursor Pol III transcripts. Unlike for mutants of Pol III transcripts (5S* and U6Δ59-72), surveillance of precursor Pol III transcripts was clearly shown to be dependent on Nrd1-Nab3 and TRAMP *in vivo*.

Chapter 6

Processing of Telomerase RNA requires components of the nuclear RNA surveillance machinery

6.1 Introduction

Chromosomal ends in eukaryotic cells are protected by a repetitive sequence T_2G_4, which forms the core of the telomere. Telomeric DNA sequences undergo attrition during DNA replication but are maintained by telomerase, an RNP containing a reverse transcriptase (Est2) amongst other proteins and *TLC1* RNA (reviewed by Chan and Blackburn, 2004). The RNA moiety of the telomerase complex serves as a template for telomeric repeat DNA synthesis by the reverse transcriptase. Despite the functional conservation of telomerase, sequences of the RNA subunits have diverged considerably. *TLC1* RNA in yeast is significantly longer than in most other organisms and carries features of an snRNA (Seto et al., 1999). The cap is post-transcriptionally converted into a 2,2,7- trimethyl guanosine cap and *TLC1* carries an Sm-binding site at the 3' terminus. Processing of the 3' end is not well understood for yeast telomerase RNA. A 3' extended, polyadenylated form of *TLC1* (herein referred to as pre-*TLC1*) has been identified (Chapon et al., 1997) but a precursor-product relationship between the polyadenylated and mature *TLC1* has not been established. Heat inactivation of a mutant form of poly(A) polymerase (Pap1) leads to loss of the poly(A)$^+$ *TLC1* signal (Chapon et al., 1997). This suggests that polyadenylation of pre-*TLC1* is mediated by the pre-mRNA cleavage and polyadenylation machinery.

6.2 Nrd1, Nab3 and Trf4 bind pre-*TLC1 in vivo*

The region downstream of the mature 3' end of *TLC1* contains Nrd1 and Nab3 consensus binding motifs. CRAC analyses identified binding sites for Nrd1, Nab3 and Trf4 on the 3' extended pre-*TLC1* RNA (Figure 6.1 A). Fragments crosslinked to Nrd1 and Nab3 included the consensus binding motifs and many sequences carried short non-templated oligo(A) tails (Figure 6.1 B). The major Trf4 binding site coincided with the previously identified cluster of poly(A) sites (Chapon et al., 1997); Figure 6.1 B), suggesting that Trf4, rather than Pap1, might be responsible for polyadenylation of the *TLC1* poly(A)$^+$ species. The few sequences found at internal *TLC1* positions did not carry any oligo(A) tails, suggesting that the recovered

oligoadenylated fragments represent processing intermediates rather than degradation intermediates, as observed for pre-snR13 (see Chapter 3). Crosslinking was seen also around position 900 in the mature RNA. In the proposed secondary structure of pre-*TLC1* (Zappulla and Cech, 2004) this region base-pairs with the 3' extension bearing the Nrd1 and Nab3 crosslinking sites (Figure 6.1 C), supporting the selective association of Nrd1, Nab3 and Trf4 with pre-*TLC1*.

Figure 6.1 (next page) Nrd1, Nab3 and Trf4 crosslink to 3' extended pre-*TLC1*
(A) Densities of high-throughput sequencing reads mapped to pre-*TLC1*.
(B) Alignment of representative high-throughput sequencing reads from the indicated IPs to pre-*TLC1*. Numbering indicates the nucleotide position with respect to nucleotide +1 in the *TLC1* gene. Arrows mark Polyadenylation sites according to (Chapon et al., 1997). Nrd1 and Nab3 consensus binding motifs are underlined in red, the Sm binding site in grey. Mismatches and deletions in sequencing reads are displayed in red.
(C) Two-dimensional structure of the terminal arm of pre-*TLC1* (Zappulla and Cech, 2004). Crosslinking sites of Nrd1 and Nab3 around positions 900 and 1200 (as in (A)) are highlighted.

A genome-wide screen for mutations that affect telomere length identified *trf4*Δ, implicating Trf4 is involved in some aspect of length control (Askree et al., 2004). I hypothesized that, guided by their consensus binding motifs, Nrd1 and Nab3 act in transcription termination of Pol II on *TLC1,* recruitment of Trf4 for oligoadenylation and Rrp6/exosome recruitment for processing of pre-*TLC1* to mature *TLC1* RNA.

6.3 The cleavage and polyadenylation machinery, Nrd1-Nab3, TRAMP4 and Rrp6 all participate in pre-*TLC1* processing

To investigate the potential role of Nrd1-Nab3 and Trf4 in the processing of pre-*TLC1*, I performed northern analyses on RNAs isolated from strains lacking Nrd1, Nab3, Rrp6, Trf4 or Trf5 (Figure 6.2 A). In WT strains mature *TLC1* and the poly(A)$^+$ pre-*TLC1* species were readily detectable and their levels had a ratio of approximately 2:1. In *trf4*Δ, *rrp6*Δ or Nab3 depleted cells this ratio changed to almost 1:1 (Figure 6.2 A and B). Pre-*TLC1:TLC1* ratios were calculated based on three independent experiments (Figure 6.2 B). Metabolic depletion of Nrd1 or loss of the poly(A) polymerase Trf5 had no effect on *TLC1* or pre-*TLC1* abundance. Depletion of Nab3 increased the levels of pre-*TLC1,* whereas mature *TLC1* was reduced, strongly indicative of a processing defect (Figure 6.2 C). In *trf4*Δ and *rrp6*Δ cells, strong accumulation of pre-*TLC1* was observed but levels of mature *TLC1* remained unchanged (Figure 6.2 C). This could reflect either loss of surveillance activity on (potentially misfolded) pre-*TLC1* or slowed processing of pre-*TLC1* in the absence of fully functional TRAMP and exosome complexes. Mature *TLC1* accumulated normally in *trf4*Δ and *rrp6*Δ strains. However, these are not conditional alleles and the cells may have adapted to utilize redundant, but less efficient, pathways to generate mature *TLC1*, causing pre-*TLC1* accumulation. Moreover, a lower rate of maturation might still allow normal *TLC1* accumulation in the slow growing *trf4*Δ and *rrp6*Δ strains.

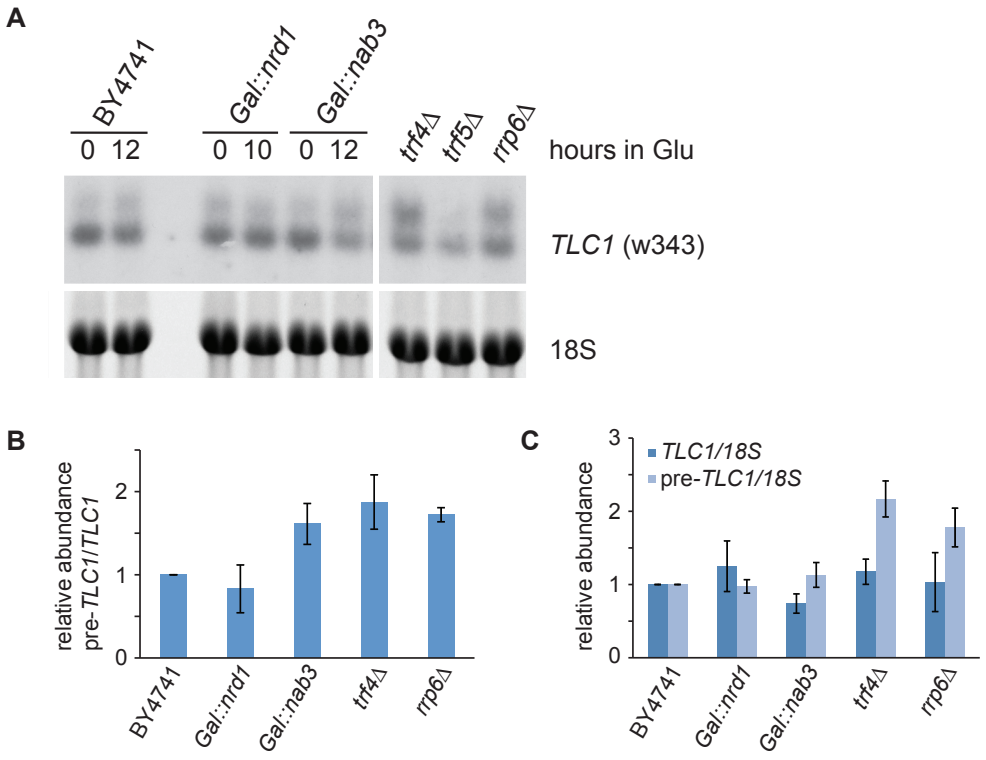

Figure 6.2 Pre-*TLC1/TLC1* abundance is altered in mutants of the surveillance machinery
(A) Northern analyses of total RNA from BY4741, P_{GAL}*::NRD1,* P_{GAL}*::NAB3, trf4*Δ, *trf5*Δ and *rrp6*Δ strains. GAL strains were pre-grown in YPGalSuc and the shifted to YPD for the indicated times prior to RNA extraction, *trf4*Δ, *trf5*Δ and *rrp6*Δ strains were grown in YPD. RNAs were separated on 1.5% BPTE agarose gel and transferred to nylon membrane. Oligo probes directed against *TLC1* are indicated and the probe number is given in brackets. Ethidium bromide staining of 18S rRNA serves as a loading control.
(B) and (C) Quantification of pre-*TLC1* and *TLC1* levels relative to each other (B) and to 18S rRNA (C). The ratio of expression levels after 12 h in glucose compared to 0 h was set to 1 for the WT. Average of three biological replicates with standard deviations.

Rather puzzling was the finding that in the absence of Trf4, levels of poly(A)$^+$ pre-*TLC1* were increased (Figures 6.2 A and 6.3), whereas poly(A)$^+$ pre-*TLC1* was reported to be undetectable after shifting a temperature sensitive *pap1-5* strain to non-permissive temperature (Chapon et al., 1997). To determine whether Pap1 is required for pre-*TLC1* accumulation in the absence of Trf4, a *pap1-5 trf4*Δ double mutant was analysed (Figure 6.3 A). In the *pap1-5* strain the pre-*TLC1* signal was lost after temperature shift. Loss of pre-*TLC1* was also observed in the *pap1-5 trf4*Δ

double mutant, suggesting that polyadenylation by Pap1 is important for pre-*TLC1* accumulation in the wild-type and in the absence of Trf4.

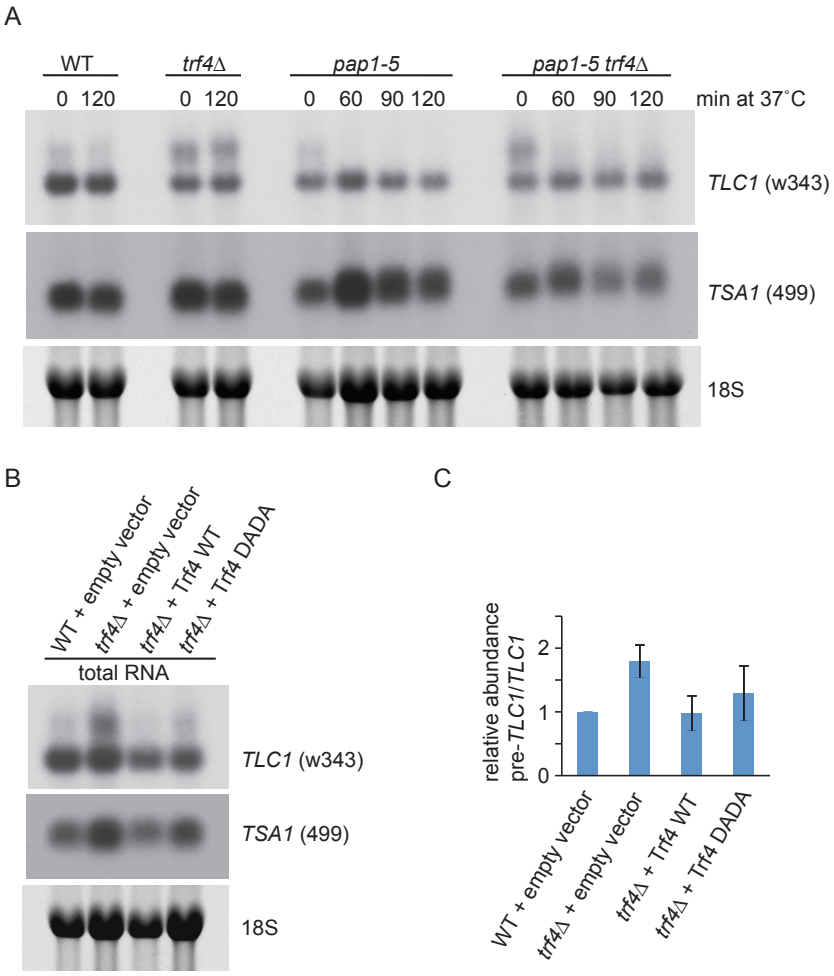

Figure 6.3 Polyadenylation activities responsible for *TLC1* processing
(A) Northern analyses of total RNA from BY4741, *trf4*Δ, *pap1-5* and *trf4*Δ *pap1-5* strains. Strains were YPD at 23˚C and shifted to 37˚C for the indicated times prior to RNA extraction. RNAs were separated on a 1.5% BPTE agarose gel and transferred to nylon membrane. Oligo probes directed against *TLC1* and *TSA1* mRNA are indicated and number of the probe is given in brackets. Ethidium bromide staining of 18S rRNA serves as a loading control.
(B) Northern analyses of total RNA from BY4741 and *trf4*Δ strains expressing plasmid borne Trf4 WT, *trf4 DADA* or an empty plasmid. RNAs were separated on a 1.5% BPTE agarose gel and transferred to nylon membrane. Oligo probes directed against *TLC1* and *TSA1* mRNA are indicated and number of the probe is given in brackets. Ethidium bromide staining of 18S rRNA serves as a loading control.
(C) Quantification of pre-*TLC1* and *TLC1* expression (from (B)) relative to each other. The ratio of expression levels was set to 1 for the WT. Average of three biological replicates with standard deviations.

Cleavage and polyadenylation of pre-mRNAs generates poly(A) tails that are coated by poly(A) binding protein (Pab1) and stabilise the RNA, whereas oligoadenylation of 3' extended pre-snoRNAs leads to exosome mediated processing (Grzechnik and Kufel, 2008). I hypothesized that oligoadenylation of pre-*TLC1* by Trf4 leads to efficient Rrp6 recruitment and processing, whereas polyadenylation of pre-*TLC1* by Pap1 leads to its accumulation. This suggested that the polyadenylation activity of Trf4 would be less important for *TLC1* processing than its interactions with the nuclear exosome. The *trf4-DADA* mutant lacks key Mg^{2+} coordinating amino acids and polyadenylation activity (Vanacova et al., 2005). In *trf4*Δ strains containing an otherwise intact TRAMP complex, expression of plasmid borne Trf4-DADA reduced pre-*TLC1* accumulation to almost WT levels (Figure 6.3 B).

6.4 Analyses of *TLC1* processing kinetics: Precursor product relationships between pre-*TLC1* and *TLC1*

Metabolic labeling of total RNA in pulse-chase experiments with [5,6-H^3] uracil or [methyl-H^3] methionine has been used for decades to study RNA processing. This is a powerful approach, particularly when combined with rapid harvesting, to give fast kinetics, and analysis by mathematical modeling (Kos and Tollervey, 2010). It is, however, applicable only to very abundant RNAs that represent a substantial fraction of total cellular RNA synthesis, such as rRNA or tRNA. To allow similar approaches to be applied to other, less abundant RNA species, 4-thiouracil was incorporated in place of [5,6-H^3] uracil as the labelling nucleotide. 4-thiouracil can be biotinylated *in vitro*, allowing newly synthesised RNA transcripts to be purified on streptavidin beads (D. Barrass, M. Kos, A. Tuck, J. Beggs and D. Tollervey, unpublished results) and analysed by a range of techniques including northern hybridisation.

I applied 4-thiouracil pulse-chase labelling and fast sampling to study *TLC1* processing (see methods Chapter 2). For this purpose a BY4741 WT strain was constructed with a genomically integrated *URA3* gene from *K. lactis* and over-expressing the uridine permease Fui1 from a plasmid. Elevated Fui1 levels increase the rate of uptake of 4-thiouracil from the media, which is limiting for incorporation,

while the *URA3* cassette allows growth in SD –URA medium prior to the pulse. Total RNA was purified from this otherwise WT yeast strain prior to 4-thiouracil addition, immediately after a 10 min pulse of 4-thiouracil incorporation (0 min time point) and at chase times up to 120 min. Recovered RNAs were further analysed by northern blotting (Figure 6.5).

In addition to pre-*TLC1* and mature *TLC1*, which were visible in steady state analyses, a further *TLC1* precursor species was detected (labelled as pre-*TLC1* (1) in Figure 6.5). The slower migrating pre-*TLC1* species (pre-*TLC1* (2) in Figure 6.5) corresponds to the mobility of the poly(A)$^+$ pre-*TLC1* detected at steady state. At the earliest chase time point pre-*TLC1* (1) was the most abundant species, but its abundance relative to pre-*TLC1* (2) and mature *TLC1* declined rapidly during the chase. This indicates that pre-*TLC1* (1) is the major initial transcript, which gives rise to pre-*TLC1* (2) and mature *TLC1*. Comparison of the processing kinetics of *TLC1* with processing of 20S to 18S rRNA suggested that pre-*TLC1* (1) is the direct precursor to *TLC1*, rather than maturation via pre-*TLC1* (2). Further experiments are under way to define the end-points and adenylation status of pre-*TLC1* (1), as well as to better determine the processing kinetics.

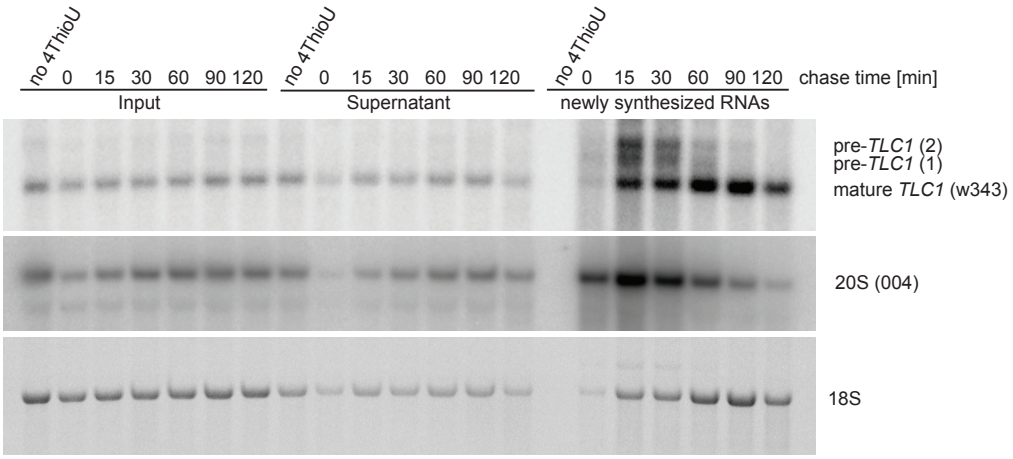

Figure 6.5 Pulse-chase analyses for *TLC1* processing
BY4741 URA3 K.L. [pRS427-Fui1] cells were grown in SD -URA -LEU. A pulse with 4-thiouracil was performed for 10 min and the cells were washed and released in SD –LEU media Nascent RNAs were purified from cells at the indicated times. 1% input, 0,75% supernatant and all of the newly synthesised RNA were analysed by northern blotting. RNAs were separated on 1.5% BPTE agarose gel and transferred to nylon membrane. Oligo probes directed against *TLC1* and 20S pre-rRNA are

indicated and the probe number is given in brackets. Ethidium bromide staining of 18S rRNA serves as a loading control.

6.5 The helicase Sen1 participates in *TLC1* transcription termination

Multiple pathways contribute to transcription termination by RNA Pol II and I wished to determine how transcription is terminated on *TLC1*. Slower migrating pre-*TLC1* species in a northern blot could represent transcription read-through products.

To test the involvement of the cleavage and polyadenylation machinery in primary transcript cleavage, ts alleles of the cleavage and polyadenylation factors Rna14, Rna15 (*rna14-1, rna15-2;* (Minvielle-Sebastia et al., 1994; Minvielle-Sebastia et al., 1991) were also tested (Figure 6.4). After shifting to 37°C, pre-*TLC1* was barely detectable in WT, *rna14-1* or *rna15-2* mutants, but there was some indication of an extended species in the *rna15-2* strain. However, these preliminary results require further confirmation.

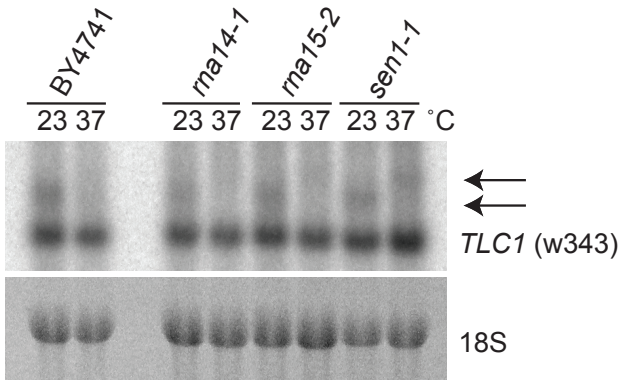

Figure 6.4 *sen1* and *rna15* mutants are defective in *TLC1* transcription termination
Northern analyses of total RNA from BY4741, *rna14-1, rna15-2* and *sen1-1* strains. Strains were grown in YPD at 23˚C and shifted to 37˚C for 30 min prior to RNA extraction. RNAs were separated on 1.5% BPTE agarose gel and transferred to nylon membrane. Oligo probes are directed against *TLC1* and pre-*TLC1* species as well as read-through products indicated by arrows. Number of the probe is given in brackets. Ethidium bromide staining of 18S rRNA serves as a loading control.

No read-through products were detected in strains depleted of Nrd1 or Nab3 (Figure 6.2 A). In contrast, a ts-lethal allele of the Nrd1-Nab3 associated helicase Sen1

(*sen1-1*) (Rasmussen and Culbertson, 1998) showed a clear shift of the pre-*TLC1* to a longer form at 37°C. This suggests that the Nrd1-Nab3-Sen1 complex participates in transcription termination on *TLC1*.

6.6 Discussion

The CRAC approach revealed the association of Nrd1, Nab3 and Trf4 with the 3' extended precursor to *TLC1* RNA (Figure 6.1). In addition, oligoadenylated fragments recovered were similar to the processing intermediates identified for snR13 (see Chapter 3 and (Grzechnik and Kufel, 2008), suggesting a processing pathway involving Nrd1-Nab3, TRAMP and the exosome. This seemed plausible given that *TLC1* has features closely resembling an snRNA and other pre-snRNAs have consensus binding motifs for Nrd1 and Nab3, which trigger their transcription termination (Steinmetz et al., 2001). Moreover, the snRNAs U1, U4 and U5 were shown to be redundantly processed by multiple exonucleases including Rrp6 and the Rex1/2/3 proteins (Allmang et al., 1999a; van Hoof et al., 2000a, b).

A model for the processing of pre-*TLC1* is presented in Fig. 6.6. Experiments with temperature sensitive mutants indicate that the Nrd1-Nab3-Sen1 complex promotes transcription termination by Pol II on *TLC1*, perhaps with a minor contribution from the mRNA cleavage and polyadenylation machinery (Figure 6.5). Further experiments with ts alleles (*nrd1-102* and *nab3-11*), which exhibit snoRNA transcription termination defects, are under way to fully elucidate the role of Nrd1-Nab3-Sen1 in transcription termination on the *TLC1* gene.

Northern analyses revealed that in mutants of Nrd1-Nab3, TRAMP4 or the nuclear exosome, processing of pre-*TLC1* is impaired (Figures 6.2 and 6.3). Expression of a polyadenylation deficient Trf4 mutant (Trf4-DADA) restored processing of pre-*TLC1* in *trf4*Δ strains. This indicates that interactions between TRAMP and the exosome/Rrp6 are necessary and sufficient to promote *TLC1* processing. Previous studies demonstrated that Pap1 is responsible for polyadenylation of pre-*TLC1* species that are detected by poly(A) selection (Chapon et al., 1997). Notably,

however, the detected, Pap1 dependent poly(A)$^+$ species probably do not correspond to the oligo(A) RNAs detected in CRAC analyses.

To better understand pre-*TLC1* processing I applied a newly developed technique for pulse-chase labelling with 4-thiouracil, which allows the isolation and analysis of newly transcribed RNAs. This provided evidence that the poly(A)$^+$ pre-*TLC1* species is neither the immediate precursor to *TLC1* nor the primary transcript. Two precursor species (pre-*TLC1* (1) and (2)) were detected, of which the slower migrating pre-*TLC1* (2) is apparently identical with poly(A)$^+$ pre-*TLC1* detected at steady state (Figure 6.5). The faster migrating pre-*TLC1* (1) species was the most abundant at early time points but rapidly declined, strongly indicating that this is the primary transcript. I predict that the TRAMP and exosome pathway normally rapidly processes this primary transcript to mature *TLC1*. A fraction of the pre-*TLC1* (1) population is polyadenylated by Pap1, generating pre-*TLC1* (2), which accumulated up to the 30 min chase time point and then slowly decreased in abundance. This shows that pre-*TLC1* (2) has a relatively long-life time but is either slowly processed to *TLC1* or degraded. It appears likely that the stability of pre-*TLC1* (2) is a consequence of polyadenylation and associated binding of Pab1. Deadenylation by Pan2-Pan3, the Ccr4-NOT complex or the exosome might be a prelude to either complete degradation or processing to *TLC1*.

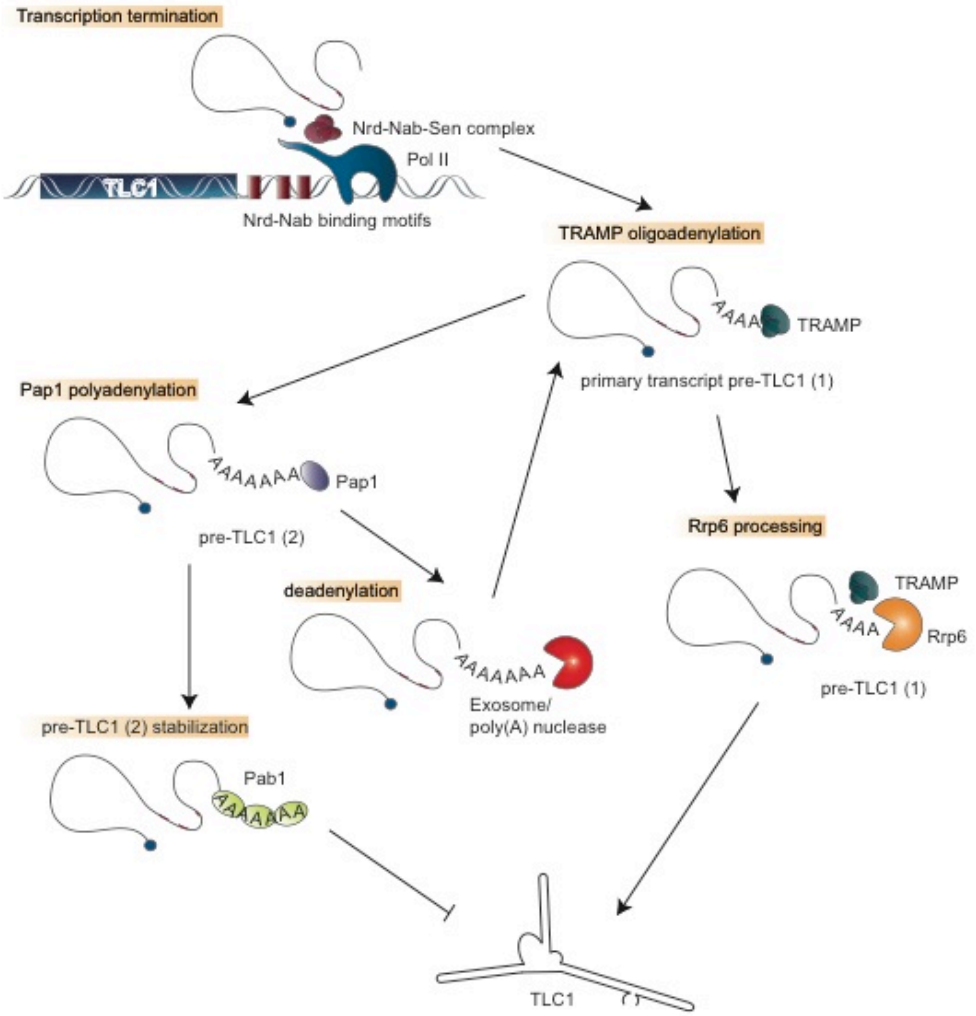

Figure 6.6 *TLC1* processing model

6.7 Applications of 4-thiouracil labelling

Analyses of *TLC1* processing following pulse-chase labelling with 4-thiouracil shows that this approach has the potential to reveal the processing kinetics of almost any RNA species. Resolution on a short time-scale allows identification of processing intermediates that are not captured otherwise. Future pulse-chase labelling in mutant backgrounds should indicate the processing steps at which specific factors are required. Specifically, this will be used to better define the roles of Trf4 and Rrp6 in *TLC1* processing. Additional applications could include 4-thiouracil labelling and deep sequencing of nascent RNAs under different growth/nutrient/mutant background conditions. This could provide information about

potential nuclear mRNA turnover pathways as was discussed in Chapter 4. The combination of 4-thiouracil pulse labelling with CRAC using RNA surveillance or processing factors should provide further insights to the intermediates bound and residence times.

Chapter 7

Overall discussions

Following their transcription RNAs undergo various steps of processing for maturation. These processes require the controlled association and dissociation of processing factors with the RNA to form a functional RNP complex. During each of these processing steps, errors are bound to occur at some frequency. In order to ensure accurate functioning of the RNP, cells have developed a very efficient RNA quality control system. It is able to recognise aberrant RNA species from several different biogenesis pathways and direct them to the exosome complex for degradation. In addition, the exosome functions in the precise processing of stable ncRNAs, including the 5.8S rRNA, snRNAs and snoRNAs

The TRAMP complex is an exosome cofactor that adds an oligo(A) tail onto aberrant RNAs, which is a mark for RNA degradation or processing. The Nrd1-Nab3 complex functions in sequence specific transcription termination of snoRNAs. During studies on degradation of CUTs it was discovered that Nrd1-Nab3 has a similar role in CUT transcription. In addition, Nrd1 interacts with the exosome *in vivo*, which underlines its participation in the nuclear RNA surveillance pathway. In the past the protein complexes, involved in surveillance, and their interactions have been extensively studied. However, not much is known about their actual RNA substrates or the mechanism by which they recognise their targets. An outstanding question is how the exosome or its cofactors are able to distinguish between processing and degradation targets.

In order to learn more about the mechanisms of nuclear RNA surveillance and potential targets I decided to apply an RNA-protein crosslinking approach, which was recently established in the lab, to factors involved in RNA degradation. This is the first genome wide study to identify targets for the surveillance machinery in a WT strain, rather than a mutant background. With this approach I was able to identify numerous known targets of the nuclear RNA surveillance machinery such as snoRNAs, CUTs, (pre-)rRNAs and certain mRNAs (Figure 3.4, 3.5, 3.7, 3.8 and 3.9). Recovery of known, very low abundance cryptic transcripts proved the high sensitivity of the crosslinking approach (*GAL10as* Figure 4.2). A general feature of many of the recovered RNAs were non-templated oligo(A) tails, demonstrating that

bona fide substrates for the nuclear RNA surveillance machinery were recovered. Bioinformatics analyses revealed that the most common A tail length on the CRAC targets was only 1-5 nt. Poly(A) tails on mRNAs associate with the poly(A) binding protein (Pab1), which both stabilizes the mRNA and stimulates translation. However, Pab1 requires a minimum binding site of ~A_{12} (Sachs et al., 1987) and is therefore not expected to bind most surveillance targets detected here, leaving them open to 3' RNase degradation. Notably, Jankowski and colleagues have observed that TRAMP preferentially adds A_4 tails when assayed *in vitro* (pers. comm.), indicating that this is an intrinsic property of the surveillance system.

In addition to the identification of known substrates for the surveillance machinery, previously unknown but anticipated RNAs were found to be associated with each of the factors Nrd1, Nab3 and Trf4. Identified targets included numerous asRNAs, mRNAs and intergenic transcripts. For a few selected RNAs (*HPF1as* and *CAF17as;* Figures 4.5 and 4.6), expression was confirmed by Northern analyses, along with stabilisation after depletion of Nrd1-Nab3 or in the absence of TRAMP or exosome components. Hit densities over annotated CUTs identified by to Xu *et al.*, revealed preferential binding of Nrd1 and Nab3 to this transcript group suggesting that direct association with the Nrd1-Nab3 complex targets RNAs to the TRAMP exosome degradation pathway.

The precursor to *TLC1*, a stable ncRNA, was also identified as a substrate for Nrd1, Nab3 and Trf4 with the CRAC approach (Figure 6.1). By northern analyses in mutants of the Nrd1-Nab3-Sen1, TRAMP4, nuclear exosome and cleavage and polyadenylation machinery, I could show that transcription termination and processing is most likely mediated by the Nrd1-Nab3-Sen1, TRAMP4 and nuclear exosome complexes (Figures 6.2, 6.3 and 6.4). Pulse-chase experiments with 4-thiouracil and newly synthesized RNA purification revealed that the previously identified poly(A)$^+$ pre-*TLC1* species is not the primary transcript and most likely not the direct precursor to *TLC1* (Figure 6.5). An additional pre-*TLC1* species was discovered, which is probably the primary *TLC1* transcript and the direct precursor to mature *TLC1*.

The most unexpected discovery was the finding that Nrd1 and Nab3 associated *in vivo* with transcripts generated by RNA Pol III (Figure 3.4 and Table 5.1). These transcripts were mainly precursors of RNA Pol III transcripts and generally carried oligo(A) tails (Figure 5.1- 5.7). Northern analysis confirmed that these, presumably defective, RNA Pol III precursor transcripts were oligoadenylated. This was shown to be dependent on recognition by Nrd1 and Nab3 and the poly(A) polymerase Trf4 (pre-*RPR1*; Figure 5.2). Furthermore, I was able to show that pre-tRNA species accumulate in the absence of Nrd1 and Nab3 (Figures 5. 5 and 5.6). This was not dependent on the association of the Nrd1-Nab3-Sen1 complex with RNA Pol II, providing further evidence that the identified Pol III RNA targets did not arise from spurious Pol II transcription.

These observations, taken together with the knowledge about nuclear RNA surveillance mechanisms, lead to a revised model for nuclear RNA degradation: Nrd1 and Nab3 can travel along with the transcribing RNA Pol II through the interaction of Nrd1 with the CTD. From this position they are able recognise consensus binding motifs in the RNA and guide transcription termination of RNA Pol II on short non-polyadenylated ncRNAs such as sn(o)RNAs and CUTs (Figure 7.1 A). The crosslinking experiments together with northern analysis revealed that the Nrd1-Nab3-Sen1 complex could also function in transcription termination of long-ncRNAs (as shown for *HPF1as*, 5kb and *TLC1*, 1.3kb). CRAC analysis of the asRNA transcripts has shown that these RNAs are frequently oligoadenylated (data not shown). Similarly, pre-*TLC1* processing was accomplished by components of the surveillance machinery, as was previously shown for snoRNAs. Therefore I conclude that these long RNA can be terminated or processed by the same pathway as the short transcripts such as CUTs and snoRNAs.

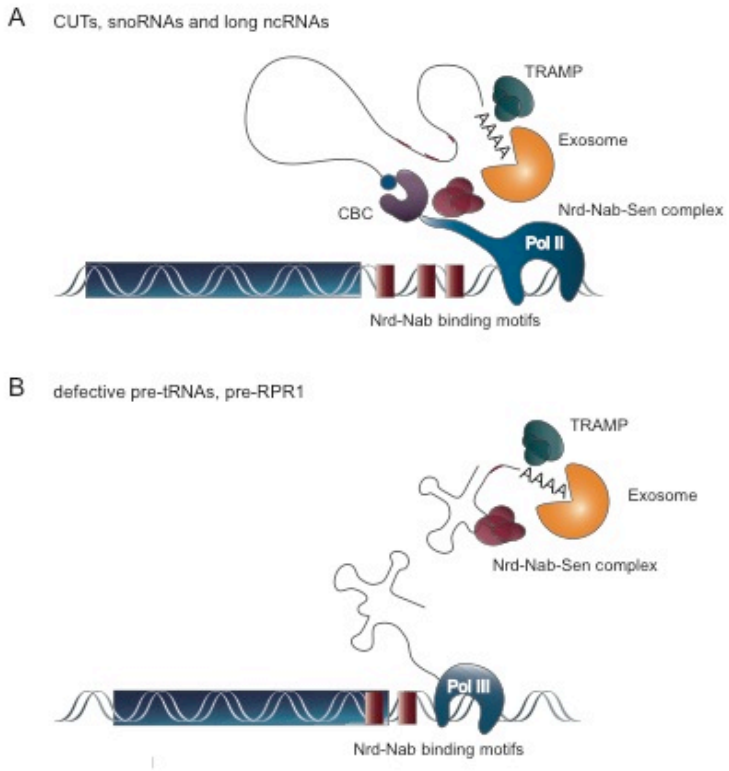

Figure 7.1 Model for nuclear RNA processing and surveillance

Considerably less was known about surveillance of RNA Pol III transcripts than Pol II derived RNAs. It was previously shown that U6 snRNA, a truncated form of 5S rRNA and undermethylated $tRNA^{iMet}$ are polyadenylated by TRAMP and degraded by the nuclear exosome (Kadaba et al., 2004; Kadaba et al., 2006; LaCava et al., 2005b; Vanacova et al., 2005). How the recognition of these substrates is mediated remained elusive, however.

I showed that Nrd1 and Nab3 are able to function independent from RNA Pol II in the recognition of (defective) RNA Pol III transcript precursors (model Figure 7.1 B), as they bind these RNAs *in vivo*. No interaction of Nrd1 or Nab3 with the RNA Pol III transcription machinery was ever observed, therefore I conclude that these factors are able to associate with their targets also post-transcriptionally. They can recognise their respective consensus binding motifs, which are carried by some, but not all, Pol III transcripts. Given that Pol III transcripts generally form very tight three-dimensional structures, I postulate that Nrd1 and Nab3 are able to recognise only those RNAs that have structural or RNP assembly defects. Once the aberrant

transcript is recognised by Nrd1-Nab3, the TRAMP complex is recruited to oligoadenylate the RNA, marking it for exosome-mediated degradation. The surveillance factors Nrd1 and Nab3 stay associated with the RNA transcripts through repetitive rounds of oligoadenylation and exosome 'chewing' until the transcript is fully degraded.

Hence, the data I present here on Nrd1, Nab3 and Trf4 target RNAs indicate that common mechanisms recognise and degrade aberrant RNAs generated by either RNA Pol II or RNA Pol III.

Bibliography

Alexandrov, A., Chernyakov, I., Gu, W., Hiley, S.L., Hughes, T.R., Grayhack, E.J., and Phizicky, E.M. (2006). Rapid tRNA Decay Can Result from Lack of Nonessential Modifications. Mol Cell *21*, 87-96.
Allmang, C., Kufel, J., Chanfreau, G., Mitchell, P., Petfalski, E., and Tollervey, D. (1999a). Functions of the exosome in rRNA, snoRNA and snRNA synthesis. EMBO J *18*, 5399-5410.
Allmang, C., Mitchell, P., Petfalski, E., and Tollervey, D. (2000). Degradation of ribosomal RNA precursors by the exosome. Nucleic Acids Res *28*, 1684-1691.
Allmang, C., Petfalski, E., Podtelejnikov, A., Mann, M., Tollervey, D., and Mitchell, P. (1999b). The yeast exosome and human PM-Scl are related complexes of 3'->5' exonucleases. Genes Dev *13*, 2148-2158.
Amberg, D.C., Goldstein, A.L., and Cole, C.N. (1992). Isolation and characterization of RAT1: an essential gene of *Saccharomyces cerevisiae* required for the efficient nucleocytoplasmic trafficking of mRNA. Genes Dev *6*, 1173-1189.
Anderson, J.S.J., and Parker, R.P. (1998). The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3' to 5' exonucleases of the exosome complex. EMBO J *17*, 1497-1506.
Arigo, J.T., Carroll, K.L., Ames, J.M., and Corden, J.L. (2006a). Regulation of yeast NRD1 expression by premature transcription termination. Mol Cell *21*, 641-651.
Arigo, J.T., Eyler, D.E., Carroll, K.L., and Corden, J.L. (2006b). Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. Mol Cell *23*, 841-851.
Auxilien, S., Crain, P.F., Trewyn, R.W., and Grosjean, H. (1996). Mechanism, specificity and general properties of the yeast enzyme catalysing the formation of inosine 34 in the anticodon of transfer RNA. J Mol Biol *262*, 437-458.
Bernstein, J., Patterson, D.N., Wilson, G.M., and Toth, E.A. (2008). Characterization of the essential activities of Saccharomyces cerevisiae Mtr4p, a 3'->5' helicase partner of the nuclear exosome. J Biol Chem *283*, 4930-4942.
Berretta, J., and Morillon, A. (2009). Pervasive transcription constitutes a new level of eukaryotic genome regulation. EMBO Rep *10*, 973-982.
Berretta, J., Pinskaya, M., and Morillon, A. (2008). A cryptic unstable transcript mediates transcriptional trans-silencing of the Ty1 retrotransposon in *S. cerevisiae*. Genes Dev *22*, 615-626.
Bertrand, E., Houser-Scott, F., Kendall, A., Singer, R.H., and Engelke, D.R. (1998). Nucleolar localization of early tRNA processing. Genes Dev *12*, 2463-2468.
Bohnsack, M.T., Martin, R., Granneman, S., Ruprecht, M., Schleiff, E., and Tollervey, D. (2009). Prp43 bound at different sites on the pre-rRNA performs distinct functions in ribosome synthesis. Mol Cell *36*, 583-592.
Bonneau, F., Basquin, J., Ebert, J., Lorentzen, E., and Conti, E. (2009). The yeast exosome functions as a macromolecular cage to channel RNA substrates for degradation. Cell *139*, 547-559.
Bosoy, D., Peng, Y., Mian, I.S., and Lue, N.F. (2003). Conserved N-terminal motifs of telomerase reverse transcriptase required for ribonucleoprotein assembly in vivo. J Biol Chem *278*, 3882-3890.
Bousquet-Antonelli, C., Presutti, C., and Tollervey, D. (2000). Identification of a regulated pathway for nuclear pre-mRNA turnover. Cell *102*, 765-775.
Box, J.A., Bunch, J.T., Tang, W., and Baumann, P. (2008). Spliceosomal cleavage generates the 3' end of telomerase RNA. Nature *456*, 910-914.

Briggs, M.W., Burkard, K.T., and Butler, J.S. (1998). Rrp6p, the yeast homologue of the human PM-Scl 100-kDa autoantigen, is essential for efficient 5.8 S rRNA 3' end formation. J Biol Chem *273*, 13255-13263.
Brown, J.T., Bai, X., and Johnson, A.W. (2000). The yeast antiviral proteins Ski2p, Ski3p, and Ski8p exist as a complex in vivo. RNA *6*, 449-457.
Burkard, K.T., and Butler, J.S. (2000). A nuclear 3'-5' exonuclease involved in mRNA degradation interacts with Poly(A) polymerase and the hnRNA protein Npl3p. Mol Cell Biol *20*, 604-616.
Cairns, B.R. (2009). The logic of chromatin architecture and remodelling at promoters. Nature *461*, 193-198.
Camblong, J., Iglesias, N., Fickentscher, C., Dieppois, G., and Stutz, F.Á. (2007). Antisense RNA stabilization induces transcriptional gene Silencing via histone deacetylation in S. cerevisiae. Cell *131*, 706-717.
Carroll, K.L., Ghirlando, R., Ames, J.M., and Corden, J.L. (2007). Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements. RNA *13*, 361-373.
Carroll, K.L., Pradhan, D.A., Granek, J.A., Clarke, N.D., and Corden, J.L. (2004). Identification of cis Elements Directing Termination of Yeast Nonpolyadenylated snoRNA Transcripts. Mol Cell Biol *24*, 6241-6252.
Chan, S.R., and Blackburn, E.H. (2004). Telomeres and telomerase. Philos Trans R Soc Lond B Biol Sci *359*, 109-121.
Chanfreau, G., Elela, S.A., Ares, M., Jr., and Guthrie, C. (1997). Alternative 3'-end processing of U5 snRNA by RNase III. Genes Dev *11*, 2741-2751.
Chanfreau, G., Legrain, P., and Jacquier, A. (1998a). Yeast RNase III as a key processing enzyme in small nucleolar RNAs metabolism. J Mol Biol *284*, 975-988.
Chanfreau, G., Rotondo, G., Legrain, P., and Jacquier, A. (1998b). Processing of a dicistronic small nucleolar RNA precursor by the RNA endonuclease Rnt1. EMBO J *17*, 3726-3737.
Chapon, C., Cech, T.R., and Zaug, A.J. (1997). Polyadenylation of telomerase RNA in budding yeast. RNA *3*, 1337-1351.
Chernyakov, I., Whipple, J.M., Kotelawala, L., Grayhack, E.J., and Phizicky, E.M. (2008). Degradation of several hypomodified mature tRNA species in Saccharomyces cerevisiae is mediated by Met22 and the 5'-3' exonucleases Rat1 and Xrn1. Genes Dev *22*, 1369-1380.
Cheung, V., Chua, G., Batada, N.N., Landry, C.R., Michnick, S.W., Hughes, T.R., and Winston, F. (2008). Chromatin- and transcription-related factors repress transcription from within coding regions throughout the Saccharomyces cerevisiae genome. PLoS Biol *6*, e277.
Ciais, D., Bohnsack, M.T., and Tollervey, D. (2008). The mRNA encoding the yeast ARE-binding protein Cth2 is generated by a novel 3' processing pathway. Nucleic Acids Res *36*, 3075-3084.
Conrad, N.K., Wilson, S.M., Steinmetz, E.J., Patturajan, M., Brow, D.A., Swanson, M.S., and Corden, J.L. (2000). A yeast heterogeneous nuclear ribonucleoprotein complex associated with RNA polymerase II. Genetics *154*, 557-571.
Copela, L.A., Fernandez, C.F., Sherrer, R.L., and Wolin, S.L. (2008). Competition between the Rex1 exonuclease and the La protein affects both Trf4p-mediated RNA quality control and pre-tRNA maturation. RNA *14*, 1214-1227.

Counter, C.M., Meyerson, M., Eaton, E.N., and Weinberg, R.A. (1997). The catalytic subunit of yeast telomerase. Proc Natl Acad Sci U S A *94*, 9202-9207.
Crick, F.H. (1966). Codon--anticodon pairing: the wobble hypothesis. J Mol Biol *19*, 548-555.
Czerwoniec, A., Dunin-Horkawicz, S., Purta, E., Kaminska, K.H., Kasprzak, J.M., Bujnicki, J.M., Grosjean, H., and Rother, K. (2009). MODOMICS: a database of RNA modification pathways. 2008 update. Nucleic Acids Res *37*, D118-121.
Dandjinou, A.T., Levesque, N., Larose, S., Lucier, J.F., Abou Elela, S., and Wellinger, R.J. (2004). A phylogenetically based secondary structure for the yeast telomerase RNA. Curr Biol *14*, 1148-1158.
Davis, C.A., and Ares, M.J. (2006). Accumulation of unstable promoter-associated transcripts upon loss of the nuclear exosome subunit Rrp6p in Saccharomyces cerevisiae. Proc Natl Acad Sci USA *103*, 3262–3267.
de la Cruz, J., Kressler, D., Tollervey, D., and Linder, P. (1998). Dob1p (Mtr4p) is a putative ATP-dependent RNA helicase required for the 3' end formation of 5.8S rRNA in Saccharomyces cerevisiae. EMBO J *17*, 1128-1140.
Dez, C., Houseley, J., and Tollervey, D. (2006). Surveillance of nuclear-restricted pre-ribosomes within a subnucleolar region of Saccharomyces cerevisiae. EMBO J *25*, 1534-1546.
Dieci, G., Fiorino, G., Castelnuovo, M., Teichmann, M., and Pagano, A. (2007). The expanding RNA polymerase III transcriptome. Trends Genet *23*, 614-622.
Doma, M.K., and Parker, R. (2007). RNA Quality Control in Eukaryotes. Cell *131*, 660-668.
Dziembowski, A., Lorentzen, E., Conti, E., and Seraphin, B. (2007). A single subunit, Dis3, is essentially responsible for yeast exosome core activity. Nat Struct Mol Biol *14*, 15-22.
El Hage, A., Koper, M., Kufel, J., and Tollervey, D. (2008). Efficient termination of transcription by RNA polymerase I requires the 5' exonuclease Rat1 in yeast. Genes Dev *22*, 1069-1081.
Erdemir, T., Bilican, B., Cagatay, T., Goding, C.R., and Yavuzer, U. (2002). Saccharomyces cerevisiae C1D is implicated in both non-homologous DNA end joining and homologous recombination. Mol Microbiol *46*, 947-957.
Frischmeyer, P.A., van Hoof, A., O'Donnell, K., Guerrerio, A.L., Parker, R., and Dietz, H.C. (2002). An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. Science *295*, 2258-2261.
Garas, M., Dichtl, B., and Keller, W. (2008). The role of the putative 3' end processing endonuclease Ysh1p in mRNA and snoRNA synthesis. RNA *14*, 2671-2684.
Geiduschek, E.P., and Kassavetis, G.A. (2001). The RNA polymerase III transcription apparatus. J Mol Biol *310*, 1-26.
Gerber, A.P., and Keller, W. (1999). An adenosine deaminase that generates inosine at the wobble position of tRNAs. Science *286*, 1146-1149.
Granneman, S., Kudla, G., Petfalski, E., and Tollervey, D. (2009). Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high throughput analysis of cDNAs. Proc Natl Acad Sci USA *106*, 9613-9818.
Greer, C.L. (1986). Assembly of a tRNA splicing complex:evidence for concerted excision and joining steps in splicing in vitro. Mol Cell Biol *6*, 635-644.

Grewal, S.I. (2010) RNAi-dependent formation of heterochromatin and its diverse functions. Curr Opin Genet Dev *20*, 134-141.
Grzechnik, P., and Kufel, J. (2008). Polyadenylation linked to transcription termination directs the processing of snoRNA precursors in yeast. Mol Cell *32*, 247-258.
Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N.R., and Altman, S. (1983). The RNA moiety of RNase P is the catalytic subunit of the enzyme. Cell *35*, 849-857.
Haracska, L., Johnson, R.E., Prakash, L., and Prakash, S. (2005). Trf4 and Trf5 proteins of Saccharomyces cerevisiae exhibit poly(A) RNA polymerase activity but no DNA polymerase activity. Mol Cell Biol *25*, 10183-10189.
Harrison, B.R., Yazgan, O., and Krebs, J.E. (2009). Life without RNAi: noncoding RNAs and their functions in Saccharomyces cerevisiae. Biochem Cell Biol *87*, 767-779.
Hentges, P., Van Driessche, B., Tafforeau, L., Vandenhaute, J., and Carr, A.M. (2005). Three novel antibiotic marker cassettes for gene disruption and marker switching in Schizosaccharomyces pombe. Yeast *22*, 1013-1019.
Hilleren, P., McCarthy, T., Rosbash, M., Parker, R., and Jensen, T.H. (2001). Quality control of mRNA 3'-end processing is linked to the nuclear exosome. Nature *413*, 538-542.
Hongay, C.F., Grisafi, P.L., Galitski, T., and Fink, G.R. (2006). Antisense transcription controls cell fate in Saccharomyces cerevisiae. Cell *127*, 735-745.
Hopper, A.K., and Phizicky, E.M. (2003). tRNA transfers to the limelight. Genes Dev *17*, 162-180.
Houalla, R., Devaux, F., Fatica, A., Kufel, J., Barrass, D., Torchet, C., and Tollervey, D. (2006). Microarray detection of novel nuclear RNA substrates for the exosome. Yeast *23*, 439-454.
Houseley, J., Kotovic, K., El Hage, A., and Tollervey, D. (2007a). Trf4 targets ncRNAs from telomeric and rDNA spacer regions and functions in rDNA copy number control. EMBO J *26*, 4996-5006.
Houseley, J., Kotovic, K., El Hage, A., and Tollervey, D. (2007b). Trf4 targets ncRNAs from telomeric and rDNA spacer regions and functions in rDNA copy number control. EMBO J *26*, 4996-5006.Houseley, J., LaCava, J., and Tollervey, D. (2006). RNA-quality control by the exosome. Nat Rev Mol Cell Biol *7*, 529-539.
Houseley, J., Rubbi, L., Grunstein, M., Tollervey, D., and Vogelauer, M. (2008). A ncRNA modulates histone modification and mRNA induction in the yeast GAL gene cluster. Mol Cell *32*, 685–695.
Houseley, J., and Tollervey, D. (2006). Yeast Trf5p is a nuclear poly(A) polymerase. EMBO Rep *7*, 205-211.
Houseley, J., and Tollervey, D. (2008). The nuclear RNA surveillance machinery: The link between ncRNAs and genome structure in budding yeast. Biochimi Biophys Acta *In Press*.
Houseley, J., and Tollervey, D. (2009). The many pathways of RNA degradation. Cell *136*, 763-776.
Huh, W.K., Falvo, J.V., Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S., and O'Shea, E.K. (2003). Global analysis of protein localization in budding yeast. Nature *425*, 686-691.

Jacquier, A. (2009). The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. Nat Rev Genet *10*, 833-844.
Johnson, A.W. (1997). Rat1p and Xrn1p are functionally interchangeable exoribonucleases that are restricted to and required in the nucleus and cytoplasm, respectively. Mol Cell Biol *17*, 6122-6130.
Kadaba, S., Krueger, A., Trice, T., Krecic, A.M., Hinnebusch, A.G., and Anderson, J. (2004). Nuclear surveillance and degradation of hypomodified initiator tRNAMet in S. cerevisiae. Genes Dev *18*, 1227-1240.
Kadaba, S., Wang, X., and Anderson, J.T. (2006). Nuclear RNA surveillance in Saccharomyces cerevisiae: Trf4p-dependent polyadenylation of nascent hypomethylated tRNA and an aberrant form of 5S rRNA. RNA *12*, 508-521.
Kaplan, C.D., Laprade, L., and Winston, F. (2003). Transcription elongation factors repress transcription initiation from cryptic sites. Science *301*, 1096-1099.
Kawauchi, J., Mischo, H., Braglia, P., Rondon, A., and Proudfoot, N.J. (2008). Budding yeast RNA polymerases I and II employ parallel mechanisms of transcriptional termination. Genes Dev *22*, 1082-1092.
Kenna, M., Stevens, A., McCammon, M., and Douglas, M.G. (1993). An essential yeast gene with homology to the exonuclease-encoding *XRN1/KEM1* gene also encodes a protein with exoribonuclease activity. Mol Cell Biol *13*, 341-350.
Kim, H., Erickson, B., Luo, W., Seward, D., Graber, J.H., Pollock, D.D., Megee, P.C., and Bentley, D.L. (2010) Gene-specific RNA polymerase II phosphorylation and the CTD code. Nat Struct Mol Biol *17*, 1279-1286.
Kim, M., Krogan, N.J., Vasiljeva, L., Rando, O.J., Nedea, E., Greenblatt, J.F., and Buratowski, S. (2004). The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. Nature *432*, 517-522.
Kim, M., Vasiljeva, L., Rando, O.J., Zhelkovsky, A., Moore, C., and Buratowski, S. (2006). Distinct pathways for snoRNA and mRNA termination. Mol Cell *24*, 723-734.
Kirmizis, A., Santos-Rosa, H., Penkett, C.J., Singer, M.A., Vermeulen, M., Mann, M., Bahler, J., Green, R.D., and Kouzarides, T. (2007). Arginine methylation at histone H3R2 controls deposition of H3K4 trimethylation. Nature *449*, 928-932.
Kobayashi, T., and Ganley, A.R. (2005). Recombination regulation by transcription-induced cohesin dissociation in rDNA repeats. Science *309*, 1581-1584.
Kos, M., and Tollervey, D. (2010) Yeast pre-rRNA processing and modification occur cotranscriptionally. Mol Cell *37*, 809-820.
Kuai, L., Das, B., and Sherman, F. (2005). A nuclear degradation pathway controls the abundance of normal mRNAs in Saccharomyces cerevisiae. Proc Natl Acad Sci USA *102*, 13962-13967.
Kufel, J., Allmang, C., Verdone, L., Beggs, J.D., and Tollervey, D. (2002). Lsm proteins are required for normal processing of pre-tRNAs and their efficient association with La (Lhp1p). Mol Cell Biol *22*, in press.
Kufel, J., and Tollervey, D. (2003). 3'-processing of yeast tRNATrp precedes 5'-processing. Rna *9*, 202-208.
LaCava, J., Houseley, J., Saveanu, C., Petfalski, E., Thompson, E., Jacquier, A., and Tollervey, D. (2005a). RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. Cell *21*, 713-724.

LaCava, J., Houseley, J., Saveanu, C., Petfalski, E., Thompson, E., Jacquier, A., and Tollervey, D. (2005b). RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. Cell *121*, 713-724.
Lafontaine, D., and Tollervey, D. (1996). One-step PCR mediated strategy for the construction of conditionally expressed and epitope tagged yeast proteins. Nucleic Acids Res *24*, 3469-3472.
Larimer, F.W., Hsu, C.L., Maupin, M.K., and Stevens, A. (1992). Characterization of the XRN1 gene encoding a 5'∅3' exoribonuclease: sequence data and analysis of disparate protein and mRNA levels of gene-disrupted yeast cells. Gene *120*, 51-57.
Lee, A., Henras, A.K., and Chanfreau, G. (2005). Multiple RNA surveillance pathways limit aberrant expression of iron uptake mRNAs and prevent iron toxicity in S. cerevisiae. Mol Cell *19*, 39-51.
Lee, J.Y., and Engelke, D.R. (1989). Partial characterization of an RNA component that copurifies with Saccharomyces cerevisiae RNase P. Mol Cell Biol *9*, 2536-2543.
Lee, J.Y., Rohlman, C.E., Molony, L.A., and Engelke, D.R. (1991). Characterization of *RPR1*, an essential gene encoding the RNA component of *Saccharomyces cerevisiae* RNase P. Mol Cell Biol *11*, 721-730.
Lee, W., Tillo, D., Bray, N., Morse, R.H., Davis, R.W., Hughes, T.R., and Nislow, C. (2007). A high-resolution atlas of nucleosome occupancy in yeast. Nat Genet *39*, 1235-1244.
Leonardi, J., Box, J.A., Bunch, J.T., and Baumann, P. (2008). TER1, the RNA subunit of fission yeast telomerase. Nat Struct Mol Biol *15*, 26-33.
Li, H., Trotta, C.R., and Abelson, J. (1998). Crystal structure and evolution of a transfer RNA splicing enzyme. Science *280*, 279-284.
Liu, Q., Greimann, J.C., and Lima, C.D. (2006). Reconstitution, activities, and structure of the eukaryotic RNA exosome. Cell *127*, 1223-1237.
Longtine, M.S., McKenzie, A., 3rd, Demarini, D.J., Shah, N.G., Wach, A., Brachat, A., Philippsen, P., and Pringle, J.R. (1998). Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. Yeast *14*, 953-961.
Luukkonen, B.G., and Seraphin, B. (1998). Construction of an in vivo-regulated U6 snRNA transcription unit as a tool to study U6 function. RNA *4*, 231-238.
Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982, revised 1989). Molecular cloning: a laboratory manual (Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press).
Martens, J.A., Laprade, L., and Winston, F. (2004). Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene. Nature *429*, 571-574.
Martens, J.A., Wu, P.Y., and Winston, F. (2005). Regulation of an intergenic transcript controls adjacent gene transcription in Saccharomyces cerevisiae. Genes Dev *19*, 2695-2704.
Martin, G., Doublie, S., and Keller, W. (2008). Determinants of substrate specificity in RNA-dependent nucleotidyl transferases. Biochim Biophys Acta *1779*, 206-216.
Martin, G., and Keller, W. (2007). RNA-specific ribonucleotidyl transferases. RNA *13*, 1834-1849.
Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Soding, J., and Cramer, P. (2010) Uniform transitions of the general RNA polymerase II transcription complex. Nat Struct Mol Biol *17*, 1272-1278.

Milligan, L., Decourty, L., Saveanu, C., Rappsilber, J., Ceulemans, H., Jacquier, A., and Tollervey, D. (2008). A yeast exosome cofactor, Mpp6, functions in RNA surveillance and in the degradation of noncoding RNA transcripts. Mol Cell Biol *28*, 5446-5457.
Milligan, L., Torchet, C., Allmang, C., Shipman, T., and Tollervey, D. (2005). A nuclear surveillance pathway for mRNAs with defective polyadenylation. Mol Cell Biol *25*, 9996-10004.
Minvielle-Sebastia, L., Preker, P.J., and Keller, W. (1994). RNA14 and RNA15 proteins as components of a yeast pre-mRNA 3'-end processing factor. Science *266*, 1702-1705.
Minvielle-Sebastia, L., Winsor, B., Bonneaud, N., and Lacroute, F. (1991). Mutations in the yeast RNA14 and RNA15 genes result in an abnormal mRNA decay rate; sequence analysis reveals an RNA-binding domain in the RNA15 protein. Mol Cell Biol *11*, 3075-3087.
Mitchell, P., Petfalski, E., Houalla, R., Podtelejnikov, A., Mann, M., and Tollervey, D. (2003). Rrp47p is an exosome-associated protein required for the 3' processing of stable RNAs. Mol Cell Biol *23*, 6982-6992.
Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M., and Tollervey, D. (1997). The exosome; a conserved eukaryotic RNA processing complex containing multiple 3'->5' exoribonuclease activities. Cell *91*, 457-466.
Mitchell, P., Petfalski, E., and Tollervey, D. (1996). The 3'-end of yeast 5.8S rRNA is generated by an exonuclease processing mechanism. Genes Dev *10*, 502-513.
Mitchell, P., and Tollervey, D. (2003). An NMD pathway in yeast involving accelerated deadenylation and exosome-mediated 3'-->5' degradation. Mol Cell *11*, 1405-1413.
Morlando, M., Ballarino, M., Greco, P., Caffarelli, E., Dichtl, B., and Bozzoni, I. (2004). Coupling between snoRNP assembly and 3' processing controls box C/D snoRNA biosynthesis in yeast. EMBO J *23*, 2392-2401.
Neil, H., Malabat, C., d/'Aubenton-Carafa, Y., Xu, Z., Steinmetz, L.M., and Jacquier, A. (2009). Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. Nature.
Parker, R., and Song, H. (2004). The enzymes and control of eukaryotic mRNA turnover. Nat Struct Mol Biol *11*, 121-127.
Petfalski, E., Dandekar, T., Henry, Y., and Tollervey, D. (1998). Processing of the precursors to small nucleolar RNAs and rRNAs requires common components. Mol Cell Biol *18*, 1181-1189.
Phizicky, E.M., and Hopper, A.K. (2010) tRNA biology charges to the front. Genes Dev *24*, 1832-1860.
Phizicky, E.M., Schwartz, R.C., and Abelson, J. (1986). *Saccharomyces cerevisiae* tRNA ligase. Purification of the protein and isolation of the structural gene. J Biol Chem *261*, 2978-2986.
Piper, P.W., Bellatin, J.A., and Lockheart, A. (1983). Altered maturation of sequence at the 3' terminus of 5S gene transcripts in a *Saccharomyces cerevisiae* mutant that lacks a RNA processing endonuclease. EMBO J *2*, 353-359.
Piper, P.W., and Stråby, K.B. (1989). Processing of transcripts of a dimeric tRNA gene in yeast uses the nuclease responsible for maturation of the 3' termini upon 5S and 37S precursor rRNAs. FEBS Lett *250*, 311-316.

Rasmussen, T.P., and Culbertson, M.R. (1998). The putative nucleic acid helicase Sen1p is required for formation and stability of termini and for maximal rates of synthesis and levels of accumulation of small nucleolar RNAs in Saccharomyces cerevisiae. Mol Cell Biol *18*, 6885-6896.
Reis, C.C., and Campbell, J.L. (2007). Contribution of Trf4/5 and the nuclear exosome to genome stability through regulation of histone mRNA levels in Saccharomyces cerevisiae. Genetics *175*, 993-1010.
Roth, K.M., Byam, J., Fang, F., and Butler, J.S. (2009). Regulation of NAB2 mRNA 3'-end formation requires the core exosome and the Trf4p component of the TRAMP complex. RNA *15*, 1045-1058.
Roth, K.M., Wolf, M.K., Rossi, M., and Butler, J.S. (2005). The nuclear exosome contributes to autogenous control of NAB2 mRNA levels. Mol Cell Biol *25*, 1577-1585.
Rouhana, L., Wang, L., Buter, N., Kwak, J.E., Schiltz, C.A., Gonzalez, T., Kelley, A.E., Landry, C.F., and Wickens, M. (2005). Vertebrate GLD2 poly(A) polymerases in the germline and the brain. RNA *11*, 1117-1130.
Sachs, A.B., Davis, R.W., and Kornberg, R.D. (1987). A single domain of yeast poly(A)-binding protein is necessary and sufficient for RNA binding and cell viability. Mol Cell Biol *7*, 3268-3276.
Sadoff, B.U., Heath-Pagliuso, S., Castano, I.B., Zhu, Y., Kieff, F.S., and Christman, M.F. (1995). Isolation of mutants of Saccharomyces cerevisiae requiring DNA topoisomerase I. Genetics *141*, 465-479.
Saitoh, S., Chabes, A., McDonald, W.H., Thelander, L., Yates, J.R., and Russell, P. (2002). Cid13 is a cytoplasmic poly(A) polymerase that regulates ribonucleotide reductase mRNA. Cell *109*, 563-573.
Schaeffer, D., Tsanova, B., A., B., Reis, F.P., Dastidar, E.G., Sanchez-Rotunno, M., Arraiano, C.M., and van Hoof, A. (2008). The exosome contains domains with specific endoribonuclease, exoribonuclease and cytoplasmic mRNA decay activities. Nat Struct Mol Biol *in press*.
Schneider, C., Anderson, J.T., and Tollervey, D. (2007). The exosome subunit Rrp44 plays a direct role in RNA substrate recognition. Mol Cell *27*, 324-331.
Schneider, C., Leung, E., Brown, J., and Tollervey, D. (2009). The N-terminal PIN domain of the exosome subunit Rrp44 harbors endonuclease activity and tethers Rrp44 to the yeast core exosome. Nucleic Acids Res *37*, 1127-1140.
Seto, A.G., Zaug, A.J., Sobel, S.G., Wolin, S.L., and Cech, T.R. (1999). Saccharomyces cerevisiae telomerase is an Sm small nuclear ribonucleoprotein particle. Nature *401*, 177-180.
Srisawat, C., Houser-Scott, F., Bertrand, E., Xiao, S., Singer, R.H., and Engelke, D.R. (2002). An active precursor in assembly of yeast nuclear ribonuclease P. RNA *8*, 1348-1360.
Steinmetz, E.J., and Brow, D.A. (1996). Repression of gene expression by an exogenous sequence element acting in concert with a heterogeneous nuclear ribonucleoprotein-like protein, Nrd1, and the putative helicase Sen1. Mol Cell Biol *16*, 6993-7003.
Steinmetz, E.J., and Brow, D.A. (1998). Control of pre-mRNA accumulation by the essential yeast protein Nrd1 requires high-affinity transcript binding and a domain implicated in RNA polymerase II association. Proc Natl Acad Sci USA *95*, 6699-6704.

Steinmetz, E.J., and Brow, D.A. (2003). Ssu72 protein mediates both poly(A)-coupled and poly(A)-independent termination of RNA polymerase II transcription. Mol Cell Biol *23*, 6339-6349.
Steinmetz, E.J., Conrad, N.K., Brow, D.A., and Corden, J.L. (2001). RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. Nature *413*, 327-331.
Symmons, M.F., Williams, M.G., Luisi, B.F., Jones, G.H., and Carpousis, A.J. (2002). Running rings around RNA: a superfamily of phosphate-dependent RNases. Trends Biochem Sci *27*, 11-18.
Takahashi, S., Araki, Y., Sakuno, T., and Katada, T. (2003). Interaction between Ski7p and Upf1p is required for nonsense-mediated 3'-to-5' mRNA decay in yeast. EMBO J *22*, 3951-3959.
Takaku, H., Minagawa, A., Takagi, M., and Nashimoto, M. (2003). A candidate prostate cancer susceptibility gene encodes tRNA 3' processing endoribonuclease. Nucleic Acids Res *31*, 2272-2278.
Thiebaut, M., Kisseleva-Romanova, E., Rougemaille, M., Boulay, J., and Libri, D. (2006). Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. Mol Cell *23*, 853-864.
Thompson, D.M., and Parker, R. (2007). Cytoplasmic decay of intergenic transcripts in Saccharomyces cerevisiae. Mol Cell Biol *27*, 92-101.
Thomson, E., and Tollervey, D. The final step in 5.8S rRNA processing is cytoplasmic in Saccharomyces cerevisiae. Mol Cell Biol *30*, 976-984.
Tong, A.H., and Boone, C. (2006). Synthetic genetic array analysis in Saccharomyces cerevisiae. Meth Mol Biol *313*, 171-192.
Torchet, C., Bousquet-Antonelli, C., Milligan, L., Thompson, E., Kufel, J., and Tollervey, D. (2002). Processing of 3' extended read-through transcripts by the exosome can generate functional mRNAs. Mol Cell *9*, 1285-1296.
Uhler, J.P., Hertel, C., and Svejstrup, J.Q. (2007). A role for noncoding transcription in activation of the yeast PHO5 gene. Proc Natl Acad Sci U S A *104*, 8011-8016.
Ule, J., Jensen, K., Mele, A., and Darnell, R.B. (2005). CLIP: a method for identifying protein-RNA interaction sites in living cells. Methods *37*, 376-386.
Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R.B. (2003). CLIP identifies Nova-regulated RNA networks in the brain. Science *302*, 1212-1215.
Ursic, D., Himmel, K.L., Gurley, K.A., Webb, F., and Culbertson, M.R. (1997). The yeast SEN1 gene is required for the processing of diverse RNA classes. Nucleic Acids Res *25*, 4778-4785.
van Hoof, A., Frischmeyer, P.A., Dietz, H.C., and Parker, R. (2002). Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. Science *295*, 2262-2264.
van Hoof, A., Lennertz, P., and Parker, R. (2000a). Three conserved members of the RNase D family have unique and overlapping functions in the processing of 5S, 5.8S, U4, U5, RNase MRP and RNase P RNAs in yeast. EMBO J *19*, 1357-1365.
van Hoof, A., Lennertz, P., and Parker, R. (2000b). Yeast exosome mutants accumulate 3'-extended polyadenylated forms of U4 small nuclear RNA and small nucleolar RNAs. Mol Cell Biol *20*, 441-452.

Vanacova, S., Wolf, J., Martin, G., Blank, D., Dettwiler, S., Friedlein, A., Langen, H., Keith, G., and Keller, W. (2005). A New Yeast Poly(A) Polymerase Complex Involved in RNA Quality Control. PLoS Biol *3*, e189.
Vasiljeva, L., and Buratowski, S. (2006). Nrd1 Interacts with the Nuclear Exosome for 3' Processing of RNA Polymerase II Transcripts. Mol Cell *21*, 239-248.
Vasiljeva, L., Kim, M., Mutschler, H., Buratowski, S., and Meinhart, A. (2008a). The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. Nat Struct Mol Biol *15*, 795-804.
Vasiljeva, L., Kim, M., Terzi, N., Soares, L.M., and Buratowski, S. (2008b). Transcription termination and RNA degradation contribute to silencing of RNA polymerase II transcription within heterochromatin. Mol Cell *29,* 313-323.
Walker, S.C., and Engelke, D.R. (2006). Ribonuclease P: the evolution of an ancient RNA enzyme. Crit Rev Biochem Mol Biol *41*, 77-102.
Wang, S.W., Toda, T., MacCallum, R., Harris, A.L., and Norbury, C. (2000a). Cid1, a fission yeast protein required for S-M checkpoint control when DNA polymerase delta or epsilon is inactivated. Mol Cell Biol *20*, 3234-3244.
Wang, X., Jia, H., Jankowsky, E., and Anderson, J.T. (2008). Degradation of hypomodified tRNA(iMet) in vivo involves RNA-dependent ATPase activity of the DExH helicase Mtr4p. RNA *14*, 107-116.
Wang, Z., Castano, I.B., Adams, C., Vu, C., Fitzhugh, D., and Christman, M.F. (2002). Structure/function analysis of the Saccharomyces cerevisiae Trf4/Pol sigma DNA polymerase. Genetics *160*, 381-391.
Wang, Z., Castano, I.B., De Las Penas, A., Adams, C., and Christman, M.F. (2000b). Pol kappa: A DNA polymerase required for sister chromatid cohesion. Science *289*, 774-779.
Webb, C.J., and Zakian, V.A. (2008). Identification and characterization of the Schizosaccharomyces pombe TER1 telomerase RNA. Nat Struct Mol Biol *15*, 34-42.
Whitehouse, I., Rando, O.J., Delrow, J., and Tsukiyama, T. (2007). Chromatin remodelling at promoters suppresses antisense transcription. Nature *450*, 1031-1035.
Wilson, S.M., Datar, K.V., Paddy, M.R., Swedlow, J.R., and Swanson, M.S. (1994). Characterization of nuclear polyadenylated RNA-binding proteins in Saccharomyces cerevisiae. J Cell Biol *127*, 1173-1184.
Winey, M., and Culbertson, M.R. (1988). Mutations affecting the tRNA-splicing endonuclease activity of Saccharomyces cerevisiae. Genetics *118*, 609-617.
Wlotzka, W., Kudla, G., Granneman, S. and Tollervey, D. (2011). The nuclear RNA polymerase II surveillance system targets polymerase III transcripts. EMBO J *in press*
Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.-C., Dufour, M.-E., Boulay, J., Régnault, B., Devaux, F., Namane, A., Séraphin, B.*, et al.* (2005). Cryptic Pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. Cell *121*, 725-737.
Xiang, S., Cooper-Morgan, A., Jiao, X., Kiledjian, M., Manley, J.L., and Tong, L. (2009). Structure and function of the 5'-->3' exoribonuclease Rat1 and its activating partner Rai1. Nature *458*, 784-788.
Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Munster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W., and Steinmetz, L.M. (2009). Bidirectional promoters generate pervasive transcription in yeast. Nature.

Yavuzer, U., Smith, G.C., Bliss, T., Werner, D., and Jackson, S.P. (1998). DNA end-independent activation of DNA-PK mediated via association with the DNA-binding protein C1D. Genes Dev *12*, 2188-2199.
Yoo, C.J., and Wolin, S.L. (1997). The yeast La protein is required for the 3' endonucleolytic cleavage that matures tRNA precursors. Cell *89*, 393-402.
Yoshihisa, T., Ohshima, C., Yunoki-Esaki, K., and Endo, T. (2007). Cytoplasmic splicing of tRNA in Saccharomyces cerevisiae. Genes Cells *12*, 285-297.
Yoshihisa, T., Yunoki-Esaki, K., Ohshima, C., Tanaka, N., and Endo, T. (2003). Possibility of cytoplasmic pre-tRNA splicing: the yeast tRNA splicing endonuclease mainly localizes on the mitochondria. Mol Biol Cell *14*, 3266-3279.
Zappulla, D.C., and Cech, T.R. (2004). Yeast telomerase RNA: a flexible scaffold for protein subunits. Proc Natl Acad Sci U S A *101*, 10024-10029.
Zhao, J., Hyman, L., and Moore, C. (1999). Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. Microbiol Mol Biol Rev *63*, 405-445.